On the use of Micro-Earthquakes to Predict Strong Ground-Motion.

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Benjamin Edwards.

April 2008.

Abstract

Since the prediction of earthquakes is, at least at present, not possible, the understanding of energy propagation from an earthquake is important in terms of mitigating structural and human losses. The physical manifestation of an earthquake is ground shaking and it is this that ultimately causes structural failure, and in the worst case scenario, loss of life. A commonly used measure of damage likelihood to a particular structure is the peak ground acceleration or velocity (PGA or PGV respectively). Typically, the prediction of PGA and PGV from a particular earthquake scenario is performed through the regression of recordings of peak ground motion. This provides predictive relations, or attenuation relations, that are valid over the magnitude range defined by the data. Unfortunately, in regions of low or moderate seismicity, the availability of recordings of strong ground-motion is unlikely. In this case the attenuation relations that are typically used have been derived from similar tectonic regions, but with higher seismicity. In doing this a bold assumption must be made: that the scaling of seismic energy is not regionally dependent. In this thesis an alternative approach is taken, where this assumption is not required. Instead, commonly available micro-earthquake data are utilised. From this data the attenuating parameters of the crust are derived along with scaling properties of the earthquake source. In order to obtain and robustly resolve a notoriously non-unique solution, a tomographic reconstruction of Q and κ is adopted. This enables the stable decoupling of geometrical decay, seismic moments, and the Brune stress drop. Using these crustal attenuation and source scaling parameters, a stochastic method is used to predict peak ground motion in terms of the 5% of critical damped response spectrum. The method is carefully tested in terms of its ability to reconstruct a synthetic dataset of known input parameters. Additionally we test the effect of varying the synthetic and inversion models in order to analyse parameter trade-off. Synthetic data are produced using a variety of methods: forward modelling of the inversion model, finite difference fault modelling and stochastic simulation. In addition bootstrap analyses are performed to estimate errors on the resultant models. It is found that the model parameters are strongly covariate, but it is shown that through the use of the tomographic reconstruction of Q, a robust solution is obtained. In addition this allows the use of a multi-dimensional Q structure as opposed to the typical homogeneous model found in other studies. The method was applied to a seismically active region: central Japan. It was shown that the results obtained using only micro-earthquake data were comparable to those obtained using strong-motion data from the same area. Additionally the response spectra for large earthquakes in Japan were modelled using a stochastic method along with the attenuation parameters derived using micro-seismicity. This was shown to be successful by comparing the model to empirical data from large earthquakes that occurred in the study region. With the knowledge that the methodology used was successful in the case of Japan it was applied to UK micro-earthquake data. Predictive ground motion relations were derived and shown to be valid over the magnitude range on record. The model was compared to other predictive relations used previously in the UK. It was shown that these alternative relations significantly over-predicted PGA and PGV for small earthquakes in the UK. Additionally the model was used to successfully predict ground motion for two of the largest UK earthquakes in the last three decades.

Chapter 1

1.1 Preface

Earthquakes are the result of a sudden, and potentially catastrophic, release of stress in the Earth's crust. The energy associated with destructive earthquakes is similar to that released by detonating megatons of TNT. This energy is radiated through the Earth by means of seismic waves. The damage caused by this is evident when we consider that over the thirty year period from 1973 to 2004 over half a million people around the world have lost their lives as a result of earthquakes (Utsu, 2002). Unfortunately however, earthquakes are an unavoidable phenomena associated with the structure and dynamics of the Earth. The reliable and precise prediction of earthquakes is, at least at present, not possible. It is clear therefore, that the understanding of the processes in the immediate aftermath of earthquakes is a fundamental requirement of mitigating both human and structural losses.

The physical manifestation of earthquakes is ground-shaking; it is this that causes damage to structures and in the worst case scenario their collapse. Modern safety-critical structures, such as chemical or nuclear facilities, are designed to withstand certain thresholds of ground motion. However, there is a significant cost-benefit trade-off in producing earthquake resistant structures. It costs much more to protect against stronger ground-motion which, at the same time, occurs less frequently. A cost effective way to reduce the seismic risk is to locate structures in areas of low seismic hazard: where the likelihood of the occurrence of strong ground-motion is lower.

In order to assess the seismic hazard of an area two main factors must be taken into account: the first is the likelihood of an earthquake occurring in a particular area over a defined time period; the second is the properties of the crust through which the radiated energy will propagate. An ideal location for a safety-critical structure would, for instance, be a structurally stable region with crustal properties that would strongly attenuate (scatter and absorb) any incident seismic energy. We can use historical records of earthquake occurrences to assess the seismicity of a region along with looking for potential sources of earthquakes, such as faults in the Earth's crust. In order to study the properties of energy attenuation in the crust we require detailed recordings of ground motion. These records are typically provided by networks of instruments deployed on a country-wide scale. For instance, the British Geological Survey currently has a network of instruments at over 140 sites across Great Britain. Vast databases of information are available, even in regions of low seismicity. This thesis is focused on obtaining these crustal attenuation properties; termed predictive ground-motion equations, or attenuation relations, using seismic databases, particularly in an absence of strong-ground motion data.

Seismic hazard is typically classified by a measure of ground motion or intensity of shaking that has a particular likelihood of not being exceeded over a given duration of time. For instance, a seismic hazard map could show the intensity of shaking that has a 90% chance of not being exceeded over 50 years. That is equivalent to a return period for that intensity of

shaking of 475 years. Considering this, we must also take into account a third factor in assessing seismic risk: the contribution of ground-motion generation. The previous assumption that a large earthquake is equivalent to high levels of ground shaking and therefore is a main contribution to seismic hazard is therefore an oversimplification. In reality, the contribution to seismic hazard must be integrated over the occurrence rate of earthquakes (e.g., Gutenburg and Richter, 1954), and the level of shaking associated with them at a given distance. Large, largest, moderate and small are obviously rather arbitrary adjectives. In this case that is required: a region such as the UK a magnitude 5 to 6 earthquake would be considered as large. Whereas in Japan a magnitude 7 earthquake would be considered large. However, the logic holds regardless of the region discussed. For consideration of short return periods (e.g., 475 years), the contribution to seismic hazard is dominated by the moderate to large sized earthquakes rather than the 'largest' possible earthquake. This is because the likelihood of the very large earthquake occurring within the short time period is very low. Comparatively, over the same duration of time the likelihood of a nearby moderate to large earthquake is much higher. On the other hand, for very long return periods (e.g., 10^6 years), for example, where considering safety-critical nuclear facilities, the contribution of the 'largest' possible earthquake to seismic hazard is considerably greater. In addition, the strict building codes for such structures means that the risk of damage is typically associated with the largest magnitude earthquakes occurring in close proximity to the structure.

1.2 Background

Understanding the generation, radiation and dissipation of seismic energy is of great importance, particularly when considering safety-critical structures. Models of this energy dissipation may be produced, termed source scaling models, predictive ground motion equations, or attenuation relations. Regardless of terminology, the models facilitate the calculation of the expected ground motion, in terms of the response of a simple harmonic oscillator. This response may be visualised over a range of natural frequencies as a response spectrum. It would be logical to question the representation of the models in terms of what is essentially an arbitrary simple harmonic oscillator (traditionally damped to 5% of critical damping). Particularly when we will go to so much trouble to remove the response of the recording instrument: leaving the exact ground response in some time derivative of displacement. The use of the response spectrum has its roots in the engineering aspect of seismology. In this field the response spectrum is frequently used as an estimation of the response of a simple structure: a building has a natural frequency, just like the simple harmonic oscillator. The response spectrum, or parts of it, can therefore be used as a measure of the likelihood of damage to a particular structure.

There are two key ways in which we can model the scaling of seismic energy. The first is the use of statistical or regression techniques on empirical data. The second is the simulation of ground motion based on the properties of the crust and the earthquake source. The former is

subject to the problem that any relationships are technically valid only within the parameter-subspace defined by the data used. The magnitude recurrence relationship of Gutenberg and Richter (1954) shows a linear relationship between the logarithm of the occurrence frequency and the magnitude of an earthquake. As reliable instrumentation networks have only been around for the last 50 or so years this means that for the larger earthquakes (relative to a particular region) it is likely that we have very few, if any, recordings. In addition, in regions of low seismicity older seismometers are were often calibrated such that recordings were on-scale for weak to moderate ground-motion: nearby (10s of km), small earthquakes ($2 < M_L < 3$) or distant (100s of km) larger ($M_L > 3$) earthquakes. Unfortunately for recordings of strong ground-motion (e.g., $M_L > 5$ at 10km) recordings were often off-scale due to the limited dynamic range of recording of the instruments. Therefore the distribution of magnitudes in our data is biased towards lower magnitudes with only distant recordings of the larger earthquakes from the historical seismicity catalogue. It is important to reiterate that in terms of hazard assessment the largest earthquakes do not necessarily contribute significantly to seismic hazard at short return periods. However, even magnitudes of $M_L = 4 - 5$ occur in the UK on average every 2 years, and $M_L > 5$ occur around every 8 years (British Geological Survey). As a result of the instrumentation limitations, data is simply not available in regions of low seismic activity for statistical regression techniques. In order to overcome this limitation, the origin of data is not typically limited to the study region (e.g., Principia Mechanica, 1982, 1988). However, this practice is reliant on the fact that the attenuation of seismic energy is not spatially variable on a regional scale. Whilst this is not as significant for the consideration of safety-critical structures The latter method is build on the theory of stochastic simulation. This is the combination of a theoretical amplitude spectrum with a random noise spectrum. The resulting time-series is then modified such that the duration of shaking is related to the earthquake's magnitude and the distance from the source. As we are really interested in the peak values of the entire time series simulated by the stochastic method, we can use random vibration theory. This very quickly estimates the peak value of a random time series relative to the root-mean-square value (see Boore, 2003, for a derivation). The main requirement of the stochastic method (or random-vibrations method) is the theoretical Fourier amplitude spectrum for an earthquake of given magnitude, stress-drop, hypocentral distance and the duration of shaking. By estimating the response ordinates in this way we hope to avoid the magnitude-validity problems associated with empirical only methods.

Determination of a model for the Fourier spectra of earthquake recordings has been achieved previously in two main ways. The first is to model the amplitude at several discrete frequencies searching for a common source model (e.g., Bay et al., 2003; Malagnini et al., 2000). The remaining path effect is allowed to vary with frequency while the variance defined by the scatter over numerous observations. This path term is then split into a geometrical decay function characterised by n segments of constant decay and a frequency dependent Q(f) term. Alternatively the whole spectrum may be modeled simultaneously (e.g.,

Scherbaum, 1990; Rietbrock, 2001; 2004). In this case the shape of the spectra is initially recovered by fitting a source model and path attenuation parameter to each event. The amplitude of the spectra are subsequently split into parameters describing geometrical spreading, seismic moments and site amplification. Each method has its own advantages and disadvantages. The modelling of discrete frequencies or peak values is arguably more vulnerable to trade-offs, for example between attenuation and the source model. In the case of the simultaneous spectral fitting there exists a further constraint in that the complete spectrum must be fit. However, this is also subject to problems, for instance if an unmodelled effect is present the fit of the whole spectrum may be be affected. In this study we follow the spectral fitting method of Scherbaum (1990) and Rietbrock (2001; 2004).

1.3 Model of Energy Propagation

The propagation of seismic energy from the earthquake source to the receiver is a complex process. As with any geophysical inversion problem we must construct a model that is suitable to the data that is available in order to obtain a robust solution. In the case of a regional network we typically have a range of data covering hypocentral distances from 10 to 100s of km. It is unlikely to obtain numerous near-source records of an earthquake. This limits the study of energy propagation to the far-field. We neglect the effects of rupture propagation: for which we would require numerous near-field recordings. Instead, we assume that the earthquake is a point source, a valid assumption for hypocentral distances much greater than the rupture length. The limitation of studying energy propagation in the far-field has implications for the applicability of results to seismic hazard analysis. The consideration of far-field induced ground motion is important particularly for non-critical structures at short return periods (e.g., 475 years). In this case the design codes are likely to be biased towards the side of cost, as opposed to design codes for safety-critical structures which are biased towards structural integrity. It is clear that the applicability of data from a relatively sparse network is not suitable for studying the very near-field maximum possible ground-motions associated with risks to safety critical structures over long return periods. However, for the consideration of moderate- to large- earthquakes at distances 10s of km from the source the data is very useful: in these circumstances structural damage still occurs (e.g., Market Rasen earthquake, UK, 27th February 2008: http://www.quakes.bgs.ac.uk/earthquakes/reports/). The results presented throughout this thesis are therefore intended to be applicable in this latter case.

It is important to understand both the nature of the earthquake source and the propagation of energy through the crust. For the transfer of energy through the crust we split the effects into two categories, that of the path through the relatively homogeneous crust, and the complex near surface site effects. These path effects include the diminution of energy through molecular level friction, the expanding wavefront, and scattering and focusing of the wave-field. In addition to the source model and the path and site effects we must also consider the response function of the recording instrument. Seismometers do not necessarily record

proportional to ground velocity over the frequency band-width required. Figure 1.1 summarises this model.

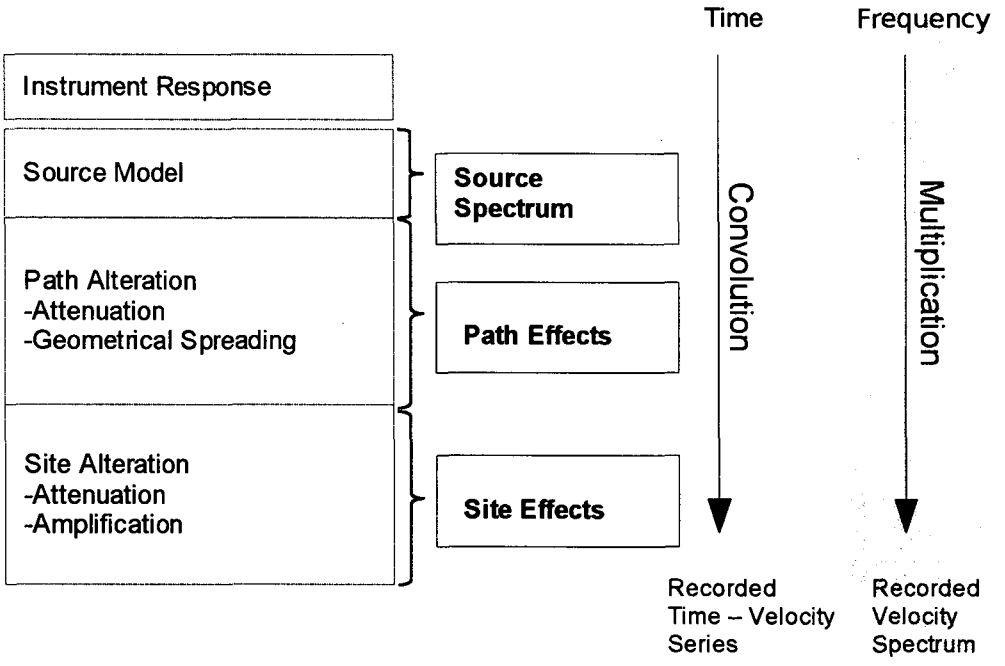

Figure 1.1: Flow chart representation of the energy propagation model. The separate parts of the model must combine (through convolution or multiplication) to form the recorded ground motion on the recording instrument.

1.3.1 Instrument Transfer Function

The seismometer response is not necessarily proportional to ground motion throughout the required passband due to the effects of damping and the natural period of the instrument. This effect must be included in the model to prevent it being mapped into the other model parameters. In the case of an electromagnetic seismometer the response is proportional down to just above the seismometer's natural frequency (Figure 1.2).

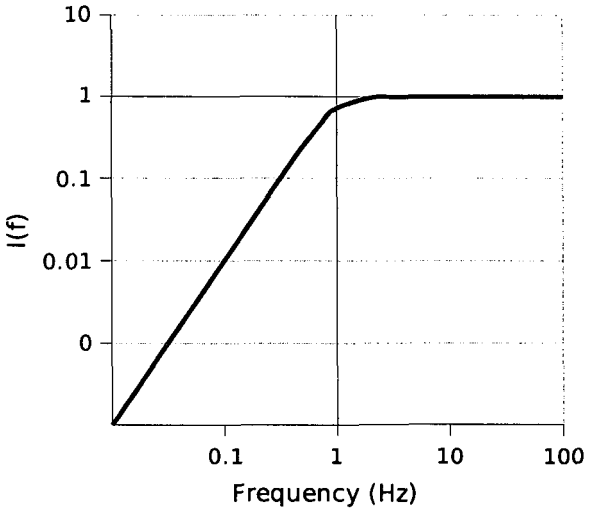

Figure 1.2: Relative response, I(f), of a velocity proportional seismometer with a 1Hz corner frequency. The response is directly proportional to velocity above around 2Hz.

The theoretical amplitude response (the modulus of the complex response) of the electromagnetic seismometer is given by:

$$I(f)=G\frac{f^2}{\sqrt{(f_0^2-f^2)^2+4h^2f^2f_0^2}} \quad , \tag{1.1}$$

where G is the generator constant of the instrument (in units of volts/ms^{-1}), f is the frequency, f_0 is the natural frequency of the instrument and h is the damping constant. For a derivation of this response see Scherbaum (2001) or Wassermann (2002). Alternatively, seismometer response may be carefully calibrated, for instance by recording the impulse response of the instrument. The response is often provided in a 'poles and zeros' format; such that the response is given by:

$$I(s)=\frac{(1-q_1s^{-1})(1-q_2s^{-1})...(1-q_Ms^{-1})}{(1-p_1s^{-1})(1-p_2s^{-1})...(1-p_Ns^{-1})} \quad , \tag{1.2}$$

where s = iω; q_i are the zeros and p_j are the poles. In addition to the mechanical response of the instrument filters are also applied during digitisation to prevent aliasing. These filters must also be included in the response function.

In order to correct for the instrument response we apply the inverse of the instrument transfer function to the recording. In the frequency domain this is achieved through the

multiplication of each datum with the reciprocal of the corresponding transfer function value ($I(f)^{-1}$).

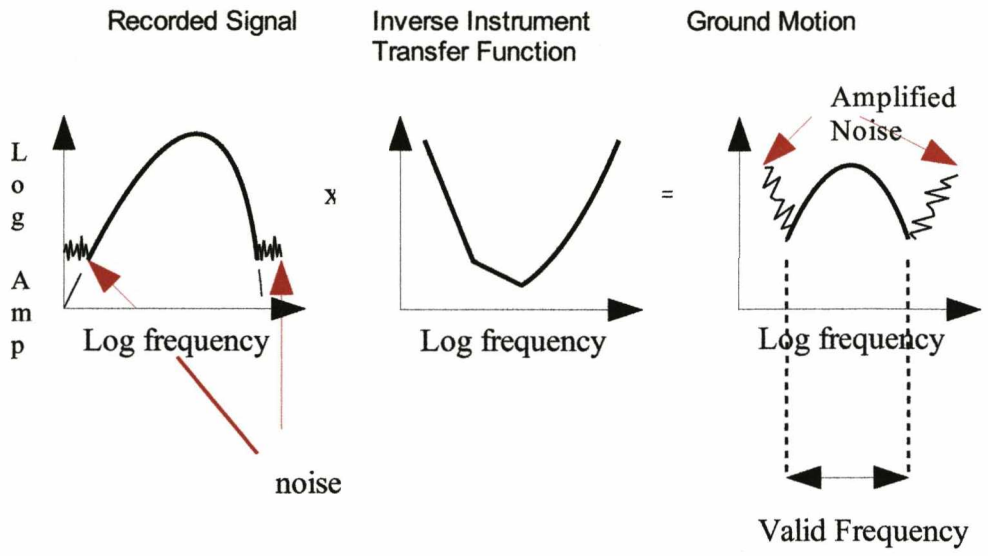

Figure 1.3: Schematic example of the removal of the instrument response from the record. Adapted from Scherbaum (2001).

Unfortunately noise occurs in the recording due to, for instance, thermal or pressure changes at the sensor, or digitisation errors such as roundoff or limited dynamic range. This is not removed by the inverse transfer function. As shown in Figure 1.3 the noise is actually increased. The valid frequency of a recording is dependent on the original noise level at the time of recording. Whilst working with this data we must avoid the noise in the record. This is achieved by taking a time-series sample from some quiet point in the record, e.g., before the first arrival. The inverse transfer function is also applied to the noise record. By comparing the noise record and the signal record we can then find the valid frequency range: where the signal is significantly greater than the noise estimate.

1.3.2 Source Model

Following Brune (1970, 1971) an earthquake source may be modelled by considering the effective stress available to accelerate the sides of a fault. The Brune (1970, 1971) model is commonly termed the ω^2 model due to its high frequency decay proportional to ω^2. This section is presented as a brief review of the model which is a key component to the method presented later on to decouple source, path and site effects. The model defines the earthquake as a "tangential stress pulse applied to the interior of the dislocation surface", Brune (1970). Fault propagation effects are neglected such that the stress pulse is applied to the entire fault surface instantaneously. Additionally, the fault surface is treated as completely reflecting to shear waves. On initiation of the rupture the center of the fault behaves as if the fault plane was infinite, and emits a pure shear stress pulse that propagates perpendicular to

the fault plane. The stress pulse time function is given by:

$$\sigma(x,t)=\sigma H(t-\frac{x}{\beta})=\mu\frac{\delta u}{\delta t}H(t-\frac{x}{\beta}) \tag{1.3}$$

where σ is the stress drop, β is the shear wave velocity and x is the distance perpendicular from fault. H(t) is the Heaviside step function defined by:

$$\begin{aligned} H(t)&=0 \quad t<0 \\ H(t)&=1 \quad t>0 \end{aligned} \tag{1.4}$$

The tangential displacement, u, is given by integration of 1.3:

$$\begin{aligned} u&=0 \qquad t<0 \\ u&=\frac{\sigma}{\mu}\beta t \quad t>0 \end{aligned} \tag{1.5}$$

The effect of diffraction is approximated by an exponential decay function with decay time of the order r_0/β, whilst the geometrical decay is accounted for by r_0/R, where r_0 is the fault radius and R is the distance from the source (hypocentral distance). The definition of r_0 is related to the point at which geometrical spreading has had no effect. For instance at $r_0=R$ the effect of geometrical spreading is described by a factor of 1. The radiation pattern is described on average by the factor $\theta_{\phi\lambda}$ = 0.55 (Boore and Boatwright , 1984). Additionally the free surface effect is described by a factor F=2.0. This is correct for horizontally polarised S waves and a good approximation for the vertical plane of motion:

$$u=F\theta_{\phi\lambda}\frac{r_0}{R}\frac{\sigma}{\mu}\beta t' e^{-2\pi f_c t'} \tag{1.6}$$

where fc is the source corner frequency and:

$$t'=t-\frac{R}{\beta} \tag{1.7}$$

such that at t'=0:

$$\frac{\delta u}{\delta t}=F\theta_{\phi\lambda}\frac{r_0}{R}\frac{\sigma}{\mu}\beta \tag{1.8}$$

The Fourier spectrum of 1.6 is:

$$E(f)=\frac{F\theta_{\phi\lambda}r_0\sigma\beta}{R\mu(f^2+f_c^2)} \quad . \tag{1.9}$$

The long-period far-field spectral density of a point source double couple is given by:

$$E_S=F\theta_{\phi\lambda}\frac{M_0}{4\pi R\rho\beta^3} \tag{1.10}$$

where M_0 is the moment of one couple of the double couple source (Keilis-Borok, 1959). If we consider the long period limit of Equation 1.9 we find that:

$$E(f\to\infty)=F\theta_{\phi\lambda}\frac{r_0}{R}\frac{\sigma\beta}{\mu f_c^2} \quad . \tag{1.11}$$

Assuming that the long-period spectral density agrees with that of a double couple source of the same moment we can combine equations 1.11 and 1.10 to obtain:

$$E_s f_c^2=F\theta_{\phi\lambda}\frac{r_0}{R}\frac{\sigma\beta}{\mu} \tag{1.12}$$

such that Equation 1.9 becomes:

$$E(f)=\frac{F\theta_{\phi\lambda}M_0}{4\pi R\rho\beta^3}\frac{f_c^2}{(f^2+f_c^2)} \quad . \tag{1.13}$$

which gives the far-field S wave spectrum. This may be generalised as:

$$E(f)=\Psi_i\cdot\frac{1}{\left(1+\left(\frac{f}{f_c}\right)^{n\gamma}\right)^{1/n}} \quad , \tag{1.14}$$

(Boatwright, 1978), where ψ_i is the long-period plateau value; f_c, is the source corner frequency, and n and γ are dependent on the source model and define the high-frequency decay at the source. For γ=2.0 and n=1.0, $E_i(f)$ is equivalent to the Brune ω^2 source model

(Brune 1970, 1971). For γ=2.0 and n=2.0, $E_i(f)$ is equivalent to the Boatwright (1978) source (Figure 1.4). The omega cubed (ω^3) source in Figure 1.4 has γ=3.0 and n=1.0.

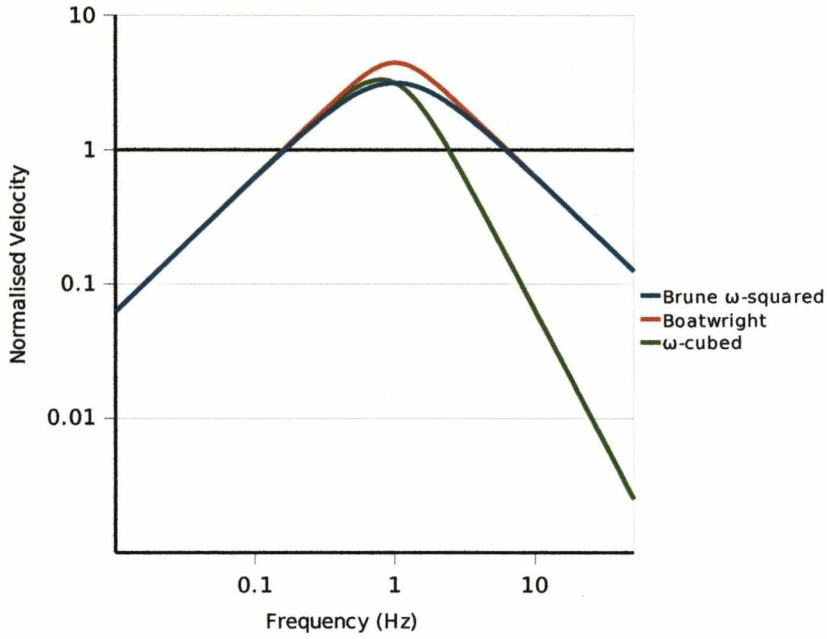

Figure 1.4: *The velocity Fourier spectra of some common earthquake source models using a 1Hz corner-frequency.*

1.3.3 Apparent Geometrical Spreading

Other processes may affect the seismic energy in a similar way to geometrical spreading (i.e., a frequency independent, distance dependent decay). This gives the appearance of super- or sub- spherical geometrical spreading ($1/R^y$ for y>1 and y<1 respectively). In addition to geometrical spreading, processes such as scattering due to heterogeneities, focusing or de-focusing of waves all contribute to this 'apparent geometrical spreading' (e.g., Atkinson and Mereu, 1992). We can describe these effects using a piecewise function of constant exponential decay across all frequencies:

$$S(r)=\begin{cases} \left(\frac{r_0}{R}\right)^{\lambda_1} & R \le R_1 \\ S(R_1)\cdot\left(\frac{R_1}{R}\right)^{\lambda_2} & R_1 \le R \le R_2 \\ \vdots & \vdots \\ S(R_n)\cdot\left(\frac{r_n}{R}\right)^{\lambda_n} & R \ge R_n \end{cases}, \qquad (1.15)$$

where R is a measure of distance to the source and r_0 is the fault radius. To comply with the assumption of a point source far-field representation of the rupture, the fault radius, r_0, is set to 1 (e.g., Boore 2003) in Equation 1.15 (e.g., note the absence of r_0 in Equation 1.10 (Keilis-Borok, 1959) and Equation 1.13 where the point source has been assumed). In the case of the point source, the definition of R is also made somewhat simpler: it is the hypocentral distance. In the case of finite fault sizes careful consideration must be taken in order to choose the correct measure of distance (i.e., at what point on the fault to take measurement from). It is also important to remain consistent in the choice of units for R: in the definition throughout this thesis SI units have been used: therefore the hypocentral distance must be measured in metres. It should also be noted that some authors (e.g., Cotton et al., in press) have found a dependence of the rate of decay on magnitude, mainly due to finite fault effects.

1.3.4 Attenuation

A further propagation effect is attenuation. The attenuation of seismic waves is due to mechanical processes at a molecular level. The degree of energy loss, and therefore attenuation, is controlled by the number of completed wave-cycles. There are therefore three parameters controlling the degree of attenuation of a seismic wave: the crustal property, the travel time and the frequency of the wave. Attenuation can be represented by the exponential function:

$$B(f)=e^{\left(-\pi f\frac{T}{Q}\right)} \tag{1.16}$$

where T is the travel time of the wave, f is the frequency of the wave, and Q is the quality factor of the propagating medium. Q^{-1} is a measure of the degree of attenuation that occurs in the medium. The longer a wave spends in the medium the more it is diminished. This is shown in Figure 1.5 for waves of different frequencies.

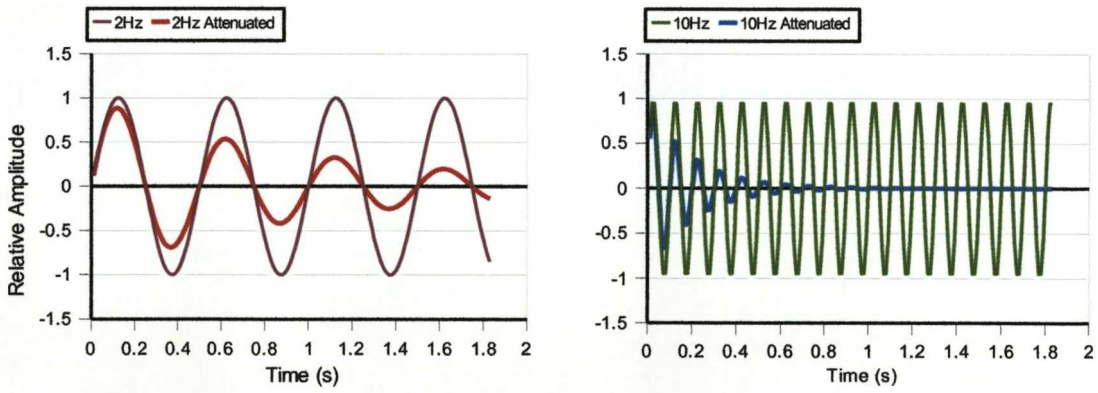

Figure 1.5: Example of attenuation of waves. Two waves of frequencies 2Hz (left panel) and 10Hz (right panel) are attenuated in the same medium over a duration of 2 seconds.

The higher the frequency of the wave, the greater it is attenuated after a particular time duration. This is shown in Figure 1.6.

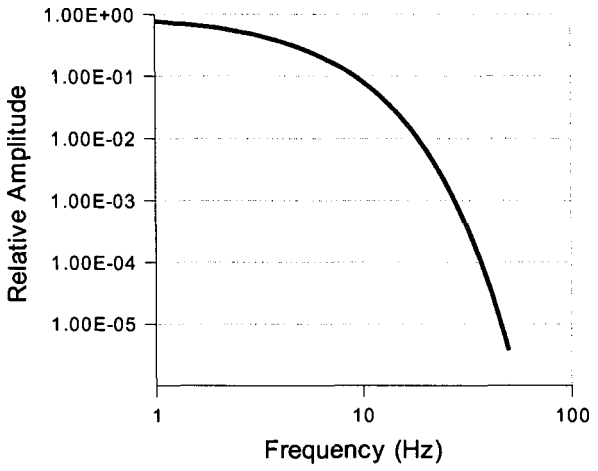

Figure 1.6: Example of the attenuation of waves of different frequencies in the same medium after a particular time duration.

For a wave of a particular frequency, the attenuation is controlled by the quality factor of the medium and the time duration of propagation. These parameters are often combined, such that attenuation is given by:

$$B(f)=e^{(-\pi f t^*)} \tag{1.17}$$

where t*=T/Q and is a measure of the degree of attenuation that the wave has undergone from source to receiver. In addition to this simple model, the quality factor may be frequency dependent:

$$Q(f)=Q_0 f^{\alpha} \tag{1.18}$$

In this case the attenuation function becomes:

$$B(f)=e^{(-\pi f^{1-\alpha} t^*)} \tag{1.19}$$

Attenuation is theoretically dependent on frequency: different wave frequencies cause different degrees of energy-loss due to different microscopic mechanisms: the relaxation spectrum for a polycrystalline material shows attenuation peaks at different frequencies due to different microscopic mechanisms. However, it may be argued that over the frequency band of interest (around 1 to 25Hz) that Q is fairly constant due to the superposition of different absorption peaks of different materials which lie at different frequencies (Liu et al., 1976).

1.3.5 Site Alteration

The complex geology at the near-surface adds further effects to the record. As with the path propagation the wave experiences attenuation. This may be represented as:

$$e^{(-\pi f^{1-\alpha} \kappa)} \tag{1.20}$$

where κ is equivalent to t* but simply differentiates the near-surface path attenuation (e.g., Anderson and Hough, 1984). Further to this, the wave field may exhibit the effects of resonance, behaving similarly to a damped harmonic oscillator (e.g., Murphy et al., 1971): the surface stratigraphy has 'natural frequencies', at which the waves seem to be amplified relative to other frequencies. The amplification of these frequencies is attributed to sharp impedance contrasts in the near surface, or due to superficial deposits of low impedance, such as soil or alluvium (eg., Beresnev and Wen, 1996). Solving the wave equation using the boundary conditions of stress and displacement at the layer boundaries can provide theoretical site response functions. For an elastic layer the response is similar to the normal modes of a damped harmonic oscillator (Murphy et al., 1971). For a viscoelastic layer the response differs in that only the first mode is present.

Combining the effects of site attenuation (Equation 1.20), resonance: 'a(f)', and mean amplification, 'A', gives the site transfer function:

$$T(f) = A \cdot a(f) \cdot e^{\left(-\pi \cdot f^{1-\alpha} \cdot \kappa\right)}, \tag{1.21}$$

Unfortunately the actual response is not always so simple, in fact the response may be also be non-linear: dependent on the magnitude of shaking. A review of this topic is provided by Beresnev and Wen (1996).

A model describing the contribution of the source, path and site effects to the Fourier spectrum of the recorded seismogram is obtained by the multiplication of Equations 1.1, 1. 14, 1.15, 1.19, and 1.21. As such, the spectral velocity, $\Omega_{ij}(f,r)$, at frequency *f* and hypocentral distance *r* is given by:

$$\Omega_{ij}(f,r) = 2\pi f E_i(f) \cdot B_{ij}(f) \cdot S_{ij}(r) \cdot T_j(f) \cdot I_j(f) \tag{1.22}$$

1.4 Data Processing and Inversion Methodology

Given that the model described in detail above describes the Fourier spectrum of a seismogram the first task is to convert the time-series database into Fourier spectra. The bilateral Laplace transform of a time series, *f(t)*, is given by:

$$F(s)=\int_{-\infty}^{\infty} e^{-st} f(t)dt \quad , \tag{1.23}$$

where, in general complex notation:

$$s=i\omega+\sigma \quad . \tag{1.24}$$

For $s=i\omega$ the bilateral Laplace transform is known as the continuous Fourier transform:

$$F(i\omega)=\int_{-\infty}^{\infty} e^{-i\omega t} f(t)dt \tag{1.25}$$

The continuous Fourier transform can be thought of as the signal in the frequency domain: the Fourier transform decomposes a time-series into a continuous spectrum of its frequency components. Along with the forward transform (Equation 1.23) there is an inverse transform:

$$f(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\omega t} F(i\omega)d\omega \quad , \tag{1.26}$$

which restores a signal from the frequency domain to the time domain. There are several properties of the continuous Fourier transform which are used in digital signal processing. One of which is the convolution theorem. This states that the convolution of two time series is a multiplication in the frequency domain:

$$f(t)*g(t)=F(i\omega).G(i\omega) \quad , \tag{1.27}$$

where * is the convolution operator. As many geophysical processes, including the model described above, are convolutions in the time-domain this makes working in the frequency domain favourable. A further property of the Fourier domain is the simplicity of differentiation and integration. Differentiation is achieved through multiplication by $i\omega$, and integration through division by $i\omega$:

$$\frac{d}{df}f(t)=\frac{d}{df}\int_{-\infty}^{\infty} e^{i\omega t} F(i\omega)dt=i\omega\int_{-\infty}^{\infty} e^{i\omega t} f(t)dt=i\omega F(i\omega) \quad . \tag{1.28}$$

In terms of the amplitude response (the modulus of the complex response) we can see that:

$$\left|\frac{d}{df}f(t)\right| = |i\omega F(i\omega)| = \omega|F(i\omega)| \tag{1.29}$$

We use this property to convert records into a common measure of ground-shaking. In this case we use velocity.

1.4.1 Tapering and Multi-Tapering

Data is recorded by seismometers at discrete time intervals (e.g., at 100Hz). The signal used in this study is therefore a windowed, finite, discrete time series. In order to obtain an estimate of the Fourier transform of discretely sampled data the Discrete Fourier Transform (DFT) can be used. In application the DFT is evaluated using the Fast Fourier Transform (FFT) algorithm, reducing the computation time by a factor of around $N/\log_2(N)$, where N is the number of samples in the input time series. A limitation of using the DFT in this case is that the time series should be periodic with periodicity over the N samples. For instance, the inverse DFT applied to the spectrum of a windowed time-series of N samples will reproduce the finite input time series but will also be periodic beyond the N input samples. Therefore at the edge of the windowed signal large discontinuities may be created (e.g., Figure 1.7).

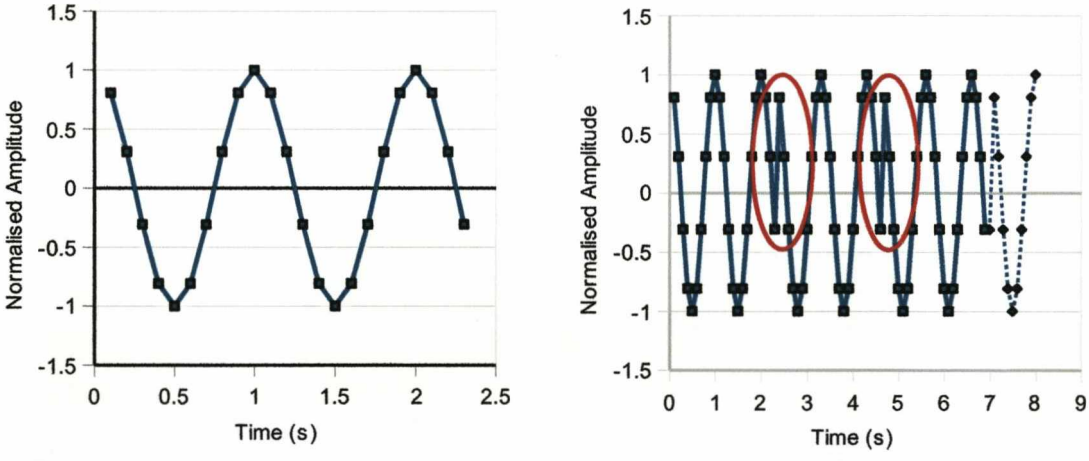

Figure 1.7: Left: a boxcar-windowed 1Hz signal. Right: Signal repeated with infinite periodicity. The discontinuity created is highlighted.

These discontinuities result in spectral leakage: the introduction of frequency content that does not really exist in the signal. To remedy this, a taper is applied to the signal before applying the Fourier transform. A typical example of a taper shape is a cosine taper. This consists of a central portion equal to one, where no change is made to the signal. Surrounding this region are two limbs of a cosine curve between 0 and 1. By applying this to the signal the sharp discontinuities are removed (e.g., Figure 1.8). This obviously removes

some information from the signal, however, it is a controlled removal and has the effect of widening and lowering the spectral peaks of the Fourier transform. For a typical seismic signal covering a wide range of frequencies this effect is averaged out and is negligible.

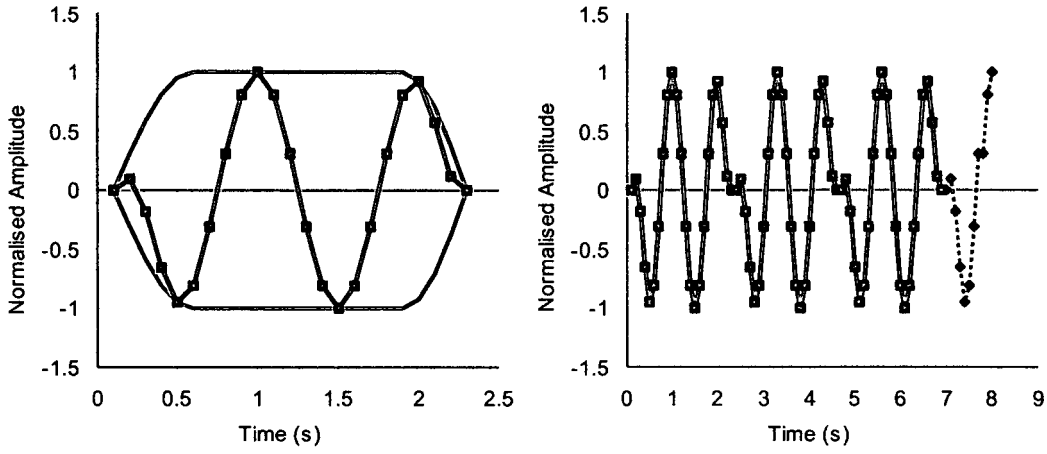

Figure 1.8: Left: The signal from Figure 1.7 has a 50% cosine taper applied to it. The cosine taper has a 50% central region equal to 1, and a limb falling to 0 on each side of the central region of width 25% of the entire taper. Right: The signal is repeated with infinite periodicity without any sharp discontinuities.

Additional spectral-leakage occurs when using the DFT due to its finite frequency resolution (i.e., discrete sampling of the Fourier Transform). As a result, frequency content of the signal between two discrete frequencies of the DFT will be distributed to the nearby resolved frequencies. Zero-padding the signal (artificially increasing the number of samples) reduces this effect.

The choice of which taper to use is subject to a trade-off between the resistance to spectral leakage and the variance of the spectral estimate (Park et al., 1987). In order to address this trade-off we can use multi-tapers: in this case the signal is multiplied by several single tapers. A Fourier transform of each of these tapered time-series is then taken before they are recombined; forming a single spectral estimate. The use of several tapers means that although some discarded by one taper, another taper will retain this data. As such the multi-taper method minimises the variance of the spectral estimate, whilst minimising the spectral leakage associated with single taper estimates. An example of a set of tapers used in the multi-taper analysis are pπ-prolate tapers e.g., Lees and Park (1995). π-prolate tapers are discrete spheroidal prolate sequences. The tapers are orthogonal, hence combining them to achieve the average spectrum does not introduce spurious correlations, as standard smoothed single-taper estimates do. In application the multi-taper achieves a smooth low variance spectrum without introducing the artifacts associated with smoothing the spectra (Figure 1.9).

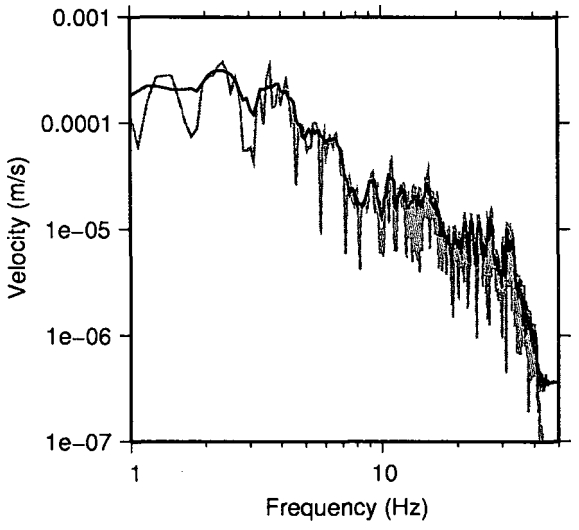

Figure 1.9: Example Fourier spectrum of a seismic record. In grey using a 20% cosine taper; and in black using a multitaper with 5 3π-prolate tapers.

1.4.2 Search Method

The inversion of Equation 1.22 is non-linear: The partial differential of 1.22 in terms of f_c is dependent on itself:

$$\begin{aligned}\frac{\partial \Omega_{ij}}{\partial f_c} &\propto \frac{d}{d f_c}\left(\frac{\Psi_i}{\left(1+\left(\frac{f}{f_c}\right)^{n\gamma}\right)^{1/n}}\right) \\ &\propto \frac{\Psi_i \gamma f\left(1+\left(\frac{f}{f_c}\right)^{n\gamma}\right)^{-\left(\frac{1}{n}+1\right)}\left(\frac{f}{f_c}\right)^{(n\gamma-1)}}{f_c^2}\end{aligned} \tag{1.30}$$

As is the partial differential of t^*:

$$\begin{aligned}\frac{\partial \Omega_{ij}}{\partial t^*} &\propto \frac{d}{dt^*} e^{\left(-\pi f^{1-\alpha} t^*\right)} \\ &\propto -\pi f^{(1-\alpha)} e^{\left(-\pi f^{1-\alpha} t^*\right)}\end{aligned} \tag{1.31}$$

Therefore an iterative approach to the problem is required. Direction-set methods were chosen due to their robustness and the flexibility of their application: such that we can simply change the constraints, model and a priori information of the inversion. Direction-set methods search a multidimensional function to find a local minimum along a search vector. There are numerous types of direction-set methods (e.g., Press et al, 1997), which differ only

by how successive search directions are chosen. A conjugate-gradient method was initially chosen: this uses the partial differentials of the function to be minimised in order to choose the new search directions. This reduces the number of iterations required compared to a method that does not use the partial differentials. However, after testing it was found that the computer-time cost of evaluating the partial differentials was more than was saved by reducing the number of iterations.

Powell's direction-set search algorithm will find the minimum of a function given an initial set of search directions: the partial differentials of the function are not required. The key to Powell's method, and indeed all direction-set methods is to find a set of N linearly independent and mutually conjugate directions (Press et al., 1997). This gives the most efficient solution, such that any search direction will not need to be repeated. We use a modified version of Powell's method (Press et al., 1997; Brent, 1973) that overcomes some of the limitations of the original method's quadratic convergence. However, the basic principal of the method, that the new search direction is the average direction of the previous N iterations, is retained.

1.4.3 Minimisation Function

Powell's method is applied to find the minimum misfit function of the form:

$$\epsilon = \sum_{f=f_{start}}^{f_{end}} \left| \Omega_{ij}^{r}(f) - \Omega_{ij}^{m}(f) \right|^2 \tag{1.32}$$

which is a least-squares (L2) minimisation where superscript *r* and *m* denote spectral values for the real data and the model respectively. The minimisation of this function finds the best values for each parameter, such that the model is as close to the data as possible in the least-squares sense. The L2 minimisation is not always ideal, particularly if there are anomalous data points, as the absolute misfit is squared, such that the inversion is heavily influenced by these outliers. An alternative misfit function is the absolute amplitude misfit (L1):

$$\epsilon = \sum_{f=f_{start}}^{f_{end}} \left| \Omega_{ij}^{r}(f) - \Omega_{ij}^{m}(f) \right| \ . \tag{1.33}$$

Alternatively, Rietbrock (1996) formulated the maximum-likelihood (ML) misfit function to be:

$$\epsilon = \sum_{f=f_{start}}^{f_{end}} \frac{\left| \Omega_{ij}^{r}(f) - \Omega_{ij}^{m}(f) \right|^2}{\Omega_{ij}^{m}(f)} \ . \tag{1.34}$$

This function considers the probability distribution of the data points: those data points with lower spectral amplitudes (which commonly lie nearer to the noise floor), are considered less reliable than data points near to the peak of the spectrum.

A further consideration is the very large range of values that make up a frequency spectrum: commonly spanning four or more orders of magnitude. If we neglect this the high frequency data, which lies at the bottom end of this range, will be neglected, in preference to fitting the high magnitude, low frequency data to a better degree. If our model is a perfect representation of the data, this would not matter as by fitting one part of the spectrum perfectly, we would fit the rest perfectly. Unfortunately, the model is never a perfect representation of real data: spurious data, unaccounted phenomena or simply noise all affect the data and it would not be feasible or possible to include all effects in the model. In order to give all parts of the spectrum a more-or-less equal influence to the misfit function we can work in the log-space. The log-space L2 function is given by:

$$\epsilon = \sum_{f=f_{start}}^{f_{end}} \left(\log(\Omega_{ij}^{r}(f)) - \log(\Omega_{ij}^{m}(f)) \right)^2 , \tag{1.35}$$

and the log-space L1 function:

$$\epsilon = \sum_{f=f_{start}}^{f_{end}} \left| \log(\Omega_{ij}^{r}(f)) - \log(\Omega_{ij}^{m}(f)) \right| . \tag{1.36}$$

The choice of minimisation function should be made after synthetic testing and bootstrap analysis in order to find the best performing function.

1.5 Inversion Procedure

1.5.1 Determining source models, attenuation and site effects.

In order to solve the minimisation function whilst minimising the trade-offs the following procedure was implemented. First, the model in Equation 1.22 is simplified, joining all the parameters affecting the amplitude of the spectra independently of frequency together. The expansion of Equation 1.22 gives:

$$\Omega_{ij}(f,r) = \Psi_i \cdot A_j \cdot S_{ij}(r) \cdot a_j(f) \cdot \frac{1}{\left(1 + \left(\frac{f}{f_c} \right)^{n\gamma} \right)^{1/n}} \cdot e^{\left(-\pi \cdot f^{1-\alpha} \cdot \left(t_{ij}^{*} + \kappa_j \right) \right)} , \tag{1.37}$$

where the first three terms are frequency independent and we collectively term the signal moment, $\hat{\Omega}_{ij}(r)$:

$$\hat{\Omega}_{ij}(r) = \Psi_i \cdot A_j \cdot S_{ij}(r) \tag{1.38}$$

It should also be noted that the zero-frequency component of the site amplification term ($a_j(f=0)$) will affect the signal moment if not equal to unity. However, in the case of this model and inversions the function $a_j(f)$ is defined such that on average we initially force $a_j(0) = 1$. As a result we expect that $a_j(0)$ is recovered from Equation 1.38 by acknowledging that $A_j \equiv a_j(0)$: the true site term is therefore given by the product $A_j\, a_j(f)$.

The minimisation function is initially solved by finding the two parameters $\hat{\Omega}_{ij}(r)$, and $(t^*+\kappa_j)$ using Powell's search algorithm. The source corner frequency is solved internally through a grid-search. The source model parameters n and γ and the frequency dependence of Q, defined by α, must be chosen prior to the inversion. Several different models may be tested, and the statistics of the fit of all data analysed to find the best source model or frequency dependence of Q. In order to constrain the inversion, the source-corner frequency is assumed to be equal for all records of a particular earthquake: the grid-search is applied to all records of one earthquake, and the best overall fit value is kept. This is not technically correct when considering directivity (e.g., Madariaga, 1976). However, for far-field observations of small earthquakes where directivity is not significant this assumption is reasonable and provides a strong constraint to the inversion. Further constraint is provided to the inversion by assuming that the path attenuation is caused by a 1D, 2D or 3D Q model; such that two identical ray-paths will be attenuated to the same degree. Similarly where two ray-paths cross they are attenuated to the same degree at that particular point (Figure 1.10).

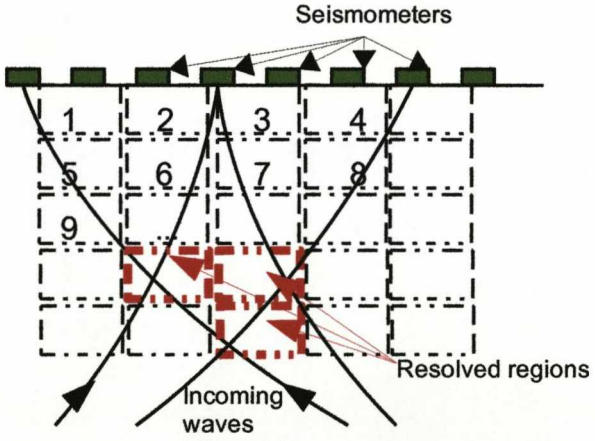

Figure 1.10: A cross section of a 3D Q model. The boxes labelled 1-9... are regions of equal Q: region 1 may have different Q to region 2 etc. The incoming waves pass through these Q regions and are attenuated. Where several paths cross (red box) we can calculate the Q value of that region.

In a way similar to velocity tomography the Q model may be solved using estimates of t* and the velocity structure of the crust. The Q model is defined such that:

$$t^*_{ij(comb)} = \int_{raypath} \frac{1}{Q(x,y,z)\cdot v(x,y,z)} ds + \kappa_j \quad , \tag{1.39}$$

where $t^*_{ij(comb)}$ is the combined path and site attenuation ($t^*+\kappa_j$). Given many $t^*_{ij(comb)}$ values we can set up a linearised damped least-squares matrix inversion. In matrix form Equation 1.39 is:

$$\boldsymbol{d} = \boldsymbol{GQ} \quad , \tag{1.40}$$

where **d** is the vector of t* data, **Q** is the Q model, and **G** is the matrix of partial differentials of the model components. An improvement to the starting model $\mathbf{Q_0}$ is obtained by solving Equation 1.40:

$$\boldsymbol{Q_1} = (\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{d} \quad . \tag{1.41}$$

The inversion is damped and iterated to find the converging solution to the Q model. We use the seismic tomography programme SIMUL2000 (Eberhart-Phillips, 1993; Thurber, 1983, 1993; Rietbrock, 2001) to produce a Q model based on the t* estimates from the Powell's search method.

The resulting Q model is used to compute the theoretical $t^*_{ij(comb)}$ values for every record in the database. The Powell's search is repeated, but this time the newly computed $t^*_{ij(comb)}$ value is fixed. By geometrically constraining the $t^*_{ij(comb)}$ values, using the Q model, we significantly reduce the trade-off between these parameters, and, by proxy the other trade-offs in the covariance matrix.

The factorial residuals of the above inversions, given by:

$$\theta_{ij}(f) = \frac{\Omega^o_{ij}(f,r)}{\Omega^m_{ij}(f,r)}, \tag{1.42}$$

where the superscripts '*o*' and '*m*' denote observed and modelled amplitudes respectively, can be used to reconstruct the frequency-dependent site function, a(f), by taking the geometrical mean of the factorial residuals at discrete frequencies over all events (i=1,2,3... n) at a specific station, j (e.g., Scherbaum, 1990; Drouet et al., 2005):

$$a_j(f)=\left(\prod_{i=1}^{n}\theta_{ij}(f)\right)^{\frac{1}{n}} . \tag{1.43}$$

1.5.2 Determining geometrical spreading, moments and site amplification.

Signal moments of each spectrum were obtained in the previous section. This is basically an amplitude parameter: altering it leads to no change in the shape of the spectrum. The signal moment is made up of the source plateau value, the geometrical spreading function and the site amplification (Equation 1.38). This is a linear problem that can be arranged in the following matrix format:

$$\begin{bmatrix} m_{11} \\ m_{12} \\ \vdots \\ m_{1J} \\ m_{21} \\ \vdots \\ m_{2J} \\ \vdots \\ m_{3J} \\ \vdots \\ m_{IJ} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 0 & 0 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & 0 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \vdots & 0 & 0 & \cdots & \vdots \\ 0 & 0 & 1 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & R'_1 & R'_2 & \cdots & R'_n & 0 & 0 & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_J \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ a_1 \\ a_2 \\ \vdots \\ a_J \end{bmatrix} , \tag{1.44}$$

where $m_{ij}=log[\hat{\Omega}_{ij}(r)]$, $\omega_i=log[\Psi_i]$, $a_j=log[A_j]$, λ_n is the constant exponent of the nth section of the apparent geometrical spreading function, and

$$R'_k=\begin{cases} log\dfrac{r_{(k-1)}}{r} & r_{k-1}<r<r_k \\ log\dfrac{r_{(k-1)}}{r_k} & r\geq r_k \end{cases} , k = 1,2,3...n. \tag{1.45}$$

$r_{k>0}$ refers to boundary distances of the kth section of the apparent geometrical spreading function and r_0 is the fault plane radius. In order to constrain the inversion we define:

$$\sum_{j=1}^{J}\log(A_j)=0 . \tag{1.46}$$

Without this constraint we could increase all magnitudes by a factor 'A' and all site amplifications by a factor of A^{-1} without affecting the misfit. It defines the parameter A_j as an

amplification relative to the network average. Equation 1.44 can be expressed as:

$$\boldsymbol{s}=\boldsymbol{G}\boldsymbol{x} \ , \tag{1.47}$$

where $\boldsymbol{s}$ is the matrix of signal moment values, $\boldsymbol{G}$ is the operator matrix and $\boldsymbol{x}$ is the model (moments, geometrical spreading exponents and site amplifications):

$$\boldsymbol{x}=(\boldsymbol{G}^T\boldsymbol{G})^{-1}\boldsymbol{G}^T\boldsymbol{s} \ . \tag{1.48}$$

Equation 1.48 is solved using a with a singular value decomposition (SVD) algorithm (Press et al., 1997; Pozo, 2004). This gives the model, $\boldsymbol{x}$, in terms of a least-squares best fit. More specifically:

$$min\left(\sum_{n=1}^{N}\left[\log(\hat{\Omega}_n(r))-s_n\right]^2=\sum_{n=1}^{N}\left[\log(\hat{\Omega}_n(r))-E[\log(\hat{\Omega}_n(r))]\right]^2=\sum_{n=1}^{N}\left(\log\left(\frac{\hat{\Omega}_n(r)}{E(\hat{\Omega}_n(r))}\right)\right)^2\right) \tag{1.49}$$

That is, we minmise the least-squares factorial residual between the real and modeled values of the signal moment in the log space. The complete procedure is summarised in the following flow chart (Figure 1.11).

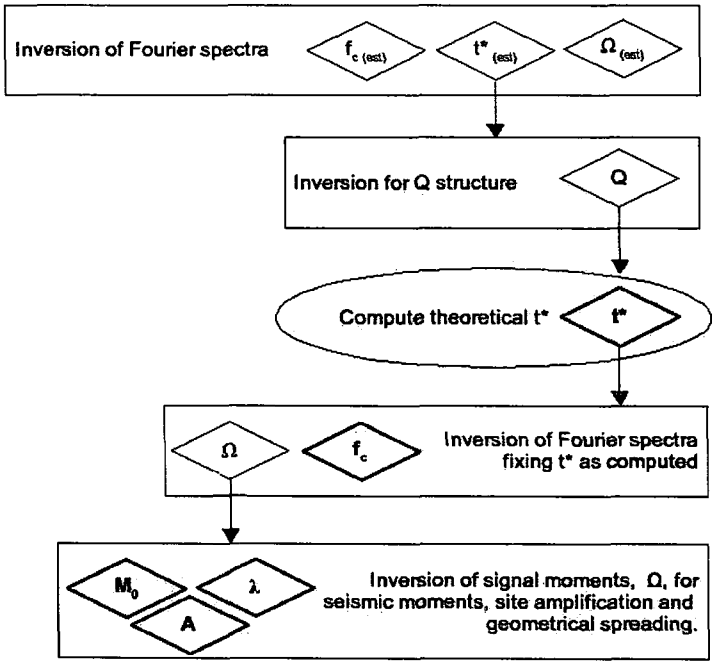

Figure 1.11: Flow diagram of basic inversion procedure. Parameters in bold are final values.

1.6 Inversion Problems: Trade-offs and Non-uniqueness

The model given by Equation 1.22 may describe the data equally well given different sets of parameters. This is known as non-uniqueness. The non-uniqueness arises due to trade- offs between the parameters in the model. An example of this is between the source model, *E(f)*, and the attenuation function, *B(f)*, or more specifically, the source corner frequency, f_c, and path attenuation, t^*, parameters. This trade-off is exemplified in Figure 1.12: different combinations of *Q* (T/t*) and f_c give very similar models. The differences are, in fact, negligible below 20Hz in this example.

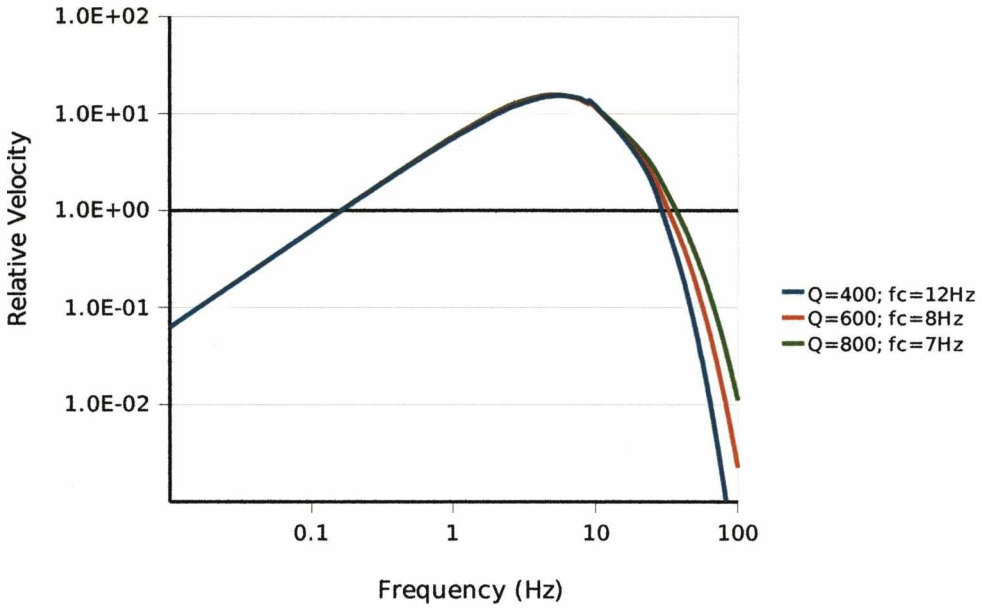

Figure 1.12: Velocity spectra for three models using different combinations of Q and f_c (see legend). All other parameters in the models are identical.

Unfortunately most seismograms are bandwidth limited to maximum frequencies of 20 to 50 Hz, to prevent anti-aliasing. In any case the noise floor limits the maximum useful frequency: higher frequencies are eclipsed by the noise. This results in difficulties in distinguishing the correct combination of t* and f_c.

Another trade-off occurs typically for high magnitude earthquakes. These typically have very low source-corner frequencies. Often the lowermost frequency data that is available is limited by the noise in the signal and the instrument type. Due to this and the low corner frequencies we tend to only observe the decaying limb of the velocity spectrum in the available bandwidth. As a result the amplitude parameters, e.g., the source plateau value and the corner frequency trade off with one another to give the same model (Figure 1.13).

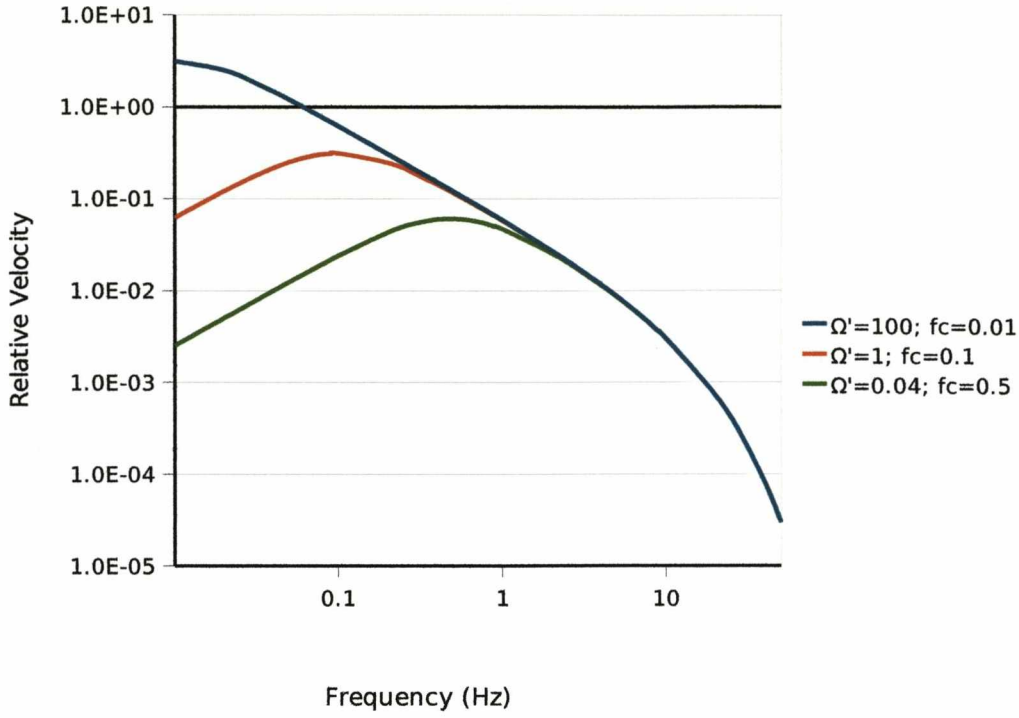

Figure 1.13: Velocity spectra for three models using different combinations of the source plateau value, Ω, and f_c (see legend). All other parameters in the models are identical.

In this plot it is apparent that if the source corner frequency is outside the instrument bandwidth it is very difficult to resolve. For instance, a common instrument bandwidth might be 1 to 25Hz. In this case it is difficult to resolve source corner frequencies below around 0.5Hz.

1.6.1 Synthetic Data

In order to gain some insight into the problem of parameter trade-off it is necessary to analyse the way in which the model parameters interact to fit synthetic datasets. By applying the methodology detailed above to a database constructed using known parameters (a synthetic dataset) we can compare the inversion results to the input parameters. The synthetic spectra were computed using the following equation:

$$\Omega_{ij}(f,r)=\frac{\Psi_i\cdot\left(\frac{R_0}{R}\right)^{\lambda_1}}{\left(1+\left(\frac{f}{f_c}\right)^{n\gamma}\right)^{1/n}}\cdot e^{\left(\frac{-\pi\cdot f\cdot t_{ij}^*}{f^{\alpha}}\right)}+P(\sigma,f)\quad . \tag{1.50}$$

This model has been simplified somewhat compared to the full inverse model (Equation 1. 37). For instance, a single segment apparent geometrical spreading function is used,

opposed to the multiple-segment function in the inverse model. Additionally site effects are made uniform: we include no resonance and no site attenuation. This allows a more straight-forward comparison between the remaining parameters in the model. A noise function, *P(σ,f)*, was included in the forward model, where σ is a unit deviate, zero mean Gaussian distribution. The function P(σ,f) is constructed such that 'coloured noise' (white noise plus Gaussian random noise) is produced:

$$P(\sigma, f) = p_1 \cdot \max(\Omega_{ij}(f,r)) \cdot (1 + p_2 \sigma) \quad , \tag{1.51}$$

where, p_1 is the fractional level of white noise compared to $\max(\Omega_{ij}(f,r))$, the peak spectral amplitude, and $p_2.p_1\max(\Omega_{ij}(f,r))$ is the standard deviation of the Gaussian noise applied to the white noise (i.e., the coloured component of the noise spectrum).

The parameters of the forward model were varied according to the test being carried out. The specific model and assumptions are given in the appropriate section. However, certain parameter values were chosen to form a standard forward model, such that:

$$\Omega_{ij}(f,r) = \frac{\Omega_i \cdot \left(\frac{1.0}{R}\right)^{1.0}}{\left(1 + \left(\frac{f}{f_c}\right)^{2.0}\right)} \cdot e^{\left(-\pi \cdot f \cdot \left(\frac{T_{ij}}{350}\right)\right)} \quad , \tag{1.52}$$

which is based on the Brune omega-squared source spectrum with frequency independent attenuation ($Q(f)=350.f^{0.0}$) and a simple, single segment apparent geometrical spreading function with an exponential decay with $\lambda=1.0$. The model is noise free unless specified to further simplify the analysis of the trade-offs. An example of the synthetic data computed using Equation 1.50 with the addition of noise using Equation 1.51 ($p_1=0.1$ and $p_2=0.2$) is shown in Figure 1.14.

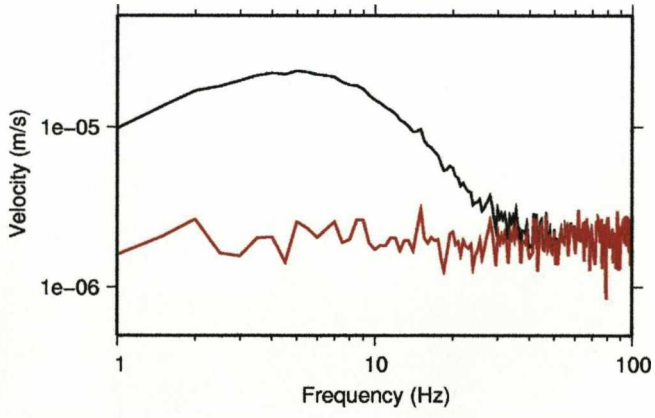

Figure 1.14: Synthetic velocity spectrum with an 8Hz source corner frequency (black line). The red line is the noise spectrum with a white noise level at 10% of the signal peak and a standard deviation of 20%.

The distribution of hypocentral distances and moment magnitudes used in the synthetic dataset are shown in Figure 1.15.

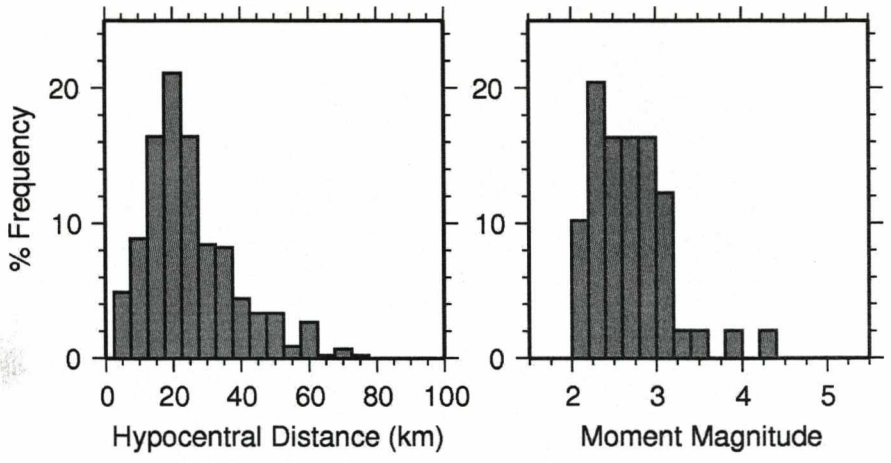

Figure 1.15: Histograms showing the range of hypocentral distances and moment magnitudes in the synthetic dataset.

The frequency of the synthetic data was bandwidth limited between 1 and 25Hz. This was based on analysis of the British Geological Survey UK database; for which the average bandwidth above a SNR of 3 over at least 5Hz was between 1.4±0.8 and 26±12Hz. The synthetic dataset was constructed to provide a dense spatial coverage of the target region. 50 earthquakes ($2 \leq M_w \leq 6$) were used, to a total of 16 stations.

The use of a homogeneous Q-structure was for simplicity, and can be likened to the analysis of a single node/block of the tomographic inversion. To reconstruct the homogeneous Q structure the harmonic mean over all observations is computed:

$$Q_0 = \frac{1}{\sum_{i=0}^{I}\sum_{j=0}^{J}\frac{1}{Q_{ij}}} = \frac{1}{\sum_{i=0}^{I}\sum_{j=0}^{J}\frac{t^*_{ij}}{T_{ij}}} \tag{1.53}$$

1.6.2 Synthetic Tests

In all cases the synthetic datasets were inverted using absolute amplitude (Equation 1.33), least-squares (Equation 1.32), maximum-likelihood (Equation 1.34) and log-space L1 and L2 (Equations 1.36 and 1.35) functions. As the dataset was constructed using known parameters, the errors in recovering these parameters were then available, and recorded. The computation of errors was performed in two stages: after the initial spectral fitting, constrained only by a common source corner frequency; then after the second inversion, constrained by both common source corner frequencies and a homogeneous Q structure. This is highlighted in Figure 1.16. The f_c grid search resolution used in all these tests was 10%.

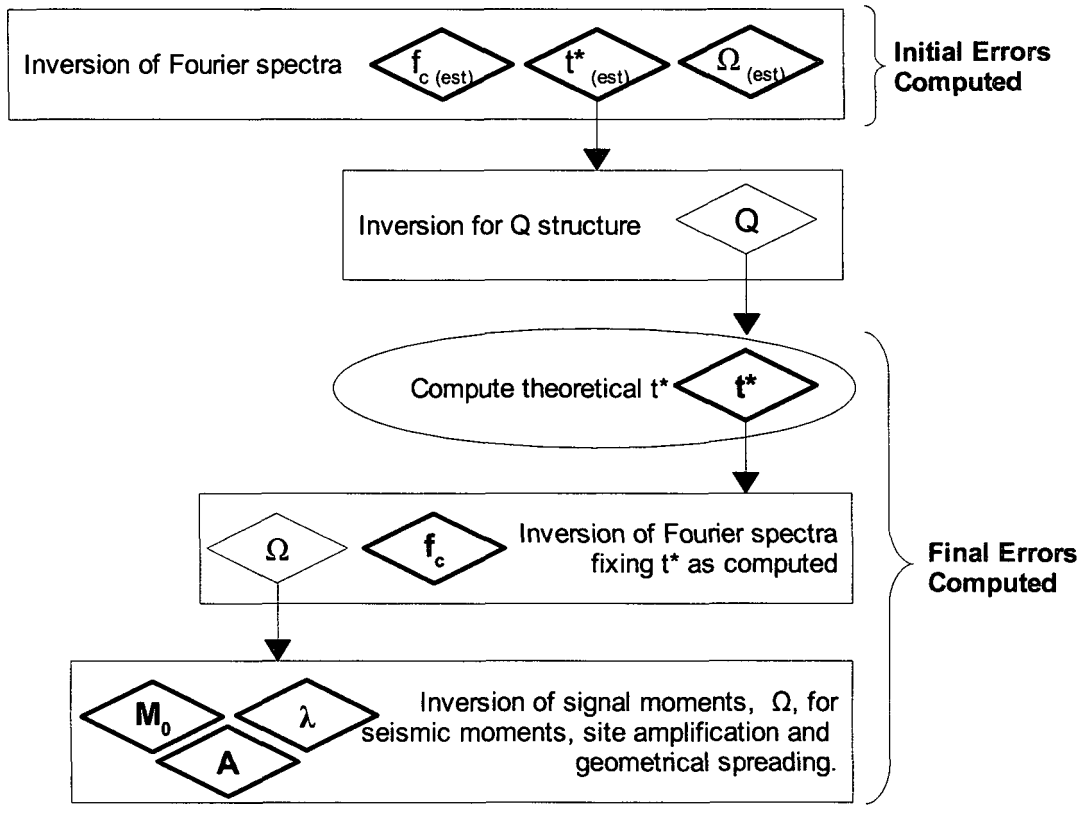

Figure 1.16: Flow chart of the inversion procedure, highlighting the points at which errors are computed.

The following tests were set up to analyse different trade-offs:

1. Incorrectly modelling the frequency dependence of Q(f).
2. Incorrectly modelling the Q structure.
3. Increasing the noise level: *P(σ,f)*.
4. Increasing the signal-to-noise ratio.
5. Incorrectly modelling the source spectrum shape.

1.6.2.1 Frequency Dependence of Q

The frequency dependence of Q, defined by:

$$Q(f)=Q_0 \cdot f^{\alpha} \tag{1.54}$$

was varied by using values of α between zero to one. No noise was included in the signal so that only the effects of incorrectly modelling the frequency dependence of Q could be analysed. When inverting the data, frequency independent Q is assumed (i.e., α_{mod}=0.0). The effects of this are an introduction of error into the other modelled parameters. As we know

the real values of those parameters this error can be measured. Figure 1.17 shows that regardless of the Q(f) function the spectrum can be fit by adjusting the source corner frequency and the signal moment.

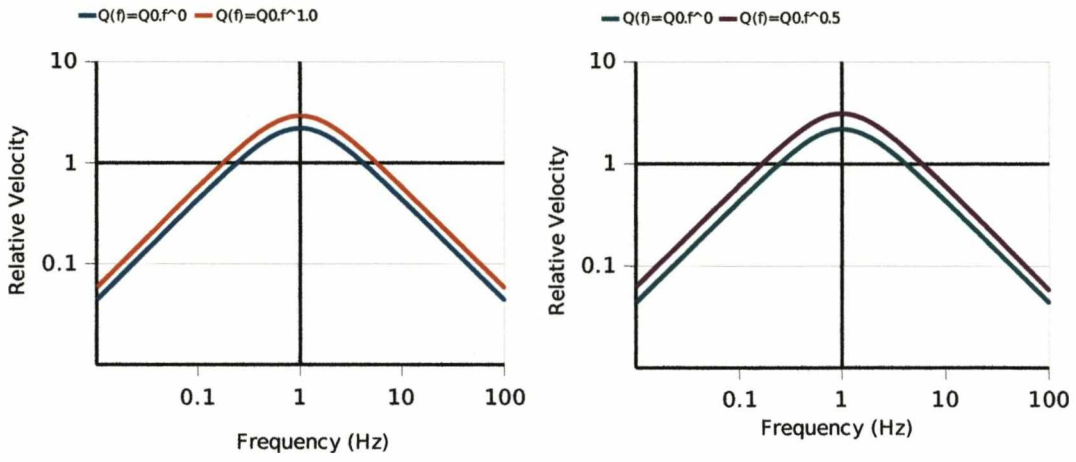

Figure 1.17: Plot of four spectral models using different Q(f) functions. Two different Q(f) = Q_0 models are overlain by two models using $\alpha \neq 0$. The signal moment (amplitude) is adjusted to vertically offset the models: otherwise they are indistinguishable.

The errors induced due to incorrectly modelling the frequency dependence of Q in the synthetic dataset are shown in two panels (Figure 1.18), one for each stage of the inversion. Each row of this figure is for a different minimisation function. Figure 1.19 shows the relative overall fit residual for each of the minimisation functions. From this we can see that the distinction between the residual fit for different α is not ideal, particularly as this is an idealised noise-free example. As α_{input} is increased to 1.0 we observe t* decreasing in an exponential fashion toward zero (Figure 1.18). The signal moments decrease to around 90% of their true value as α_{input} increases to 0.5; they then regain some, but not all, of that loss as α_{input} increases further. Overall f_c remains largely unaffected by the incorrect modelling of Q(f), particularly when you consider the grid-search resolution was only 10% of successive values (e.g., 10,11,12.1,13.3Hz...). However, there is up to 20% scatter (to one standard deviation) on these f_c error estimates. A significant reduction of this scatter is achieved after fixing the Q structure; down to just a few percent.

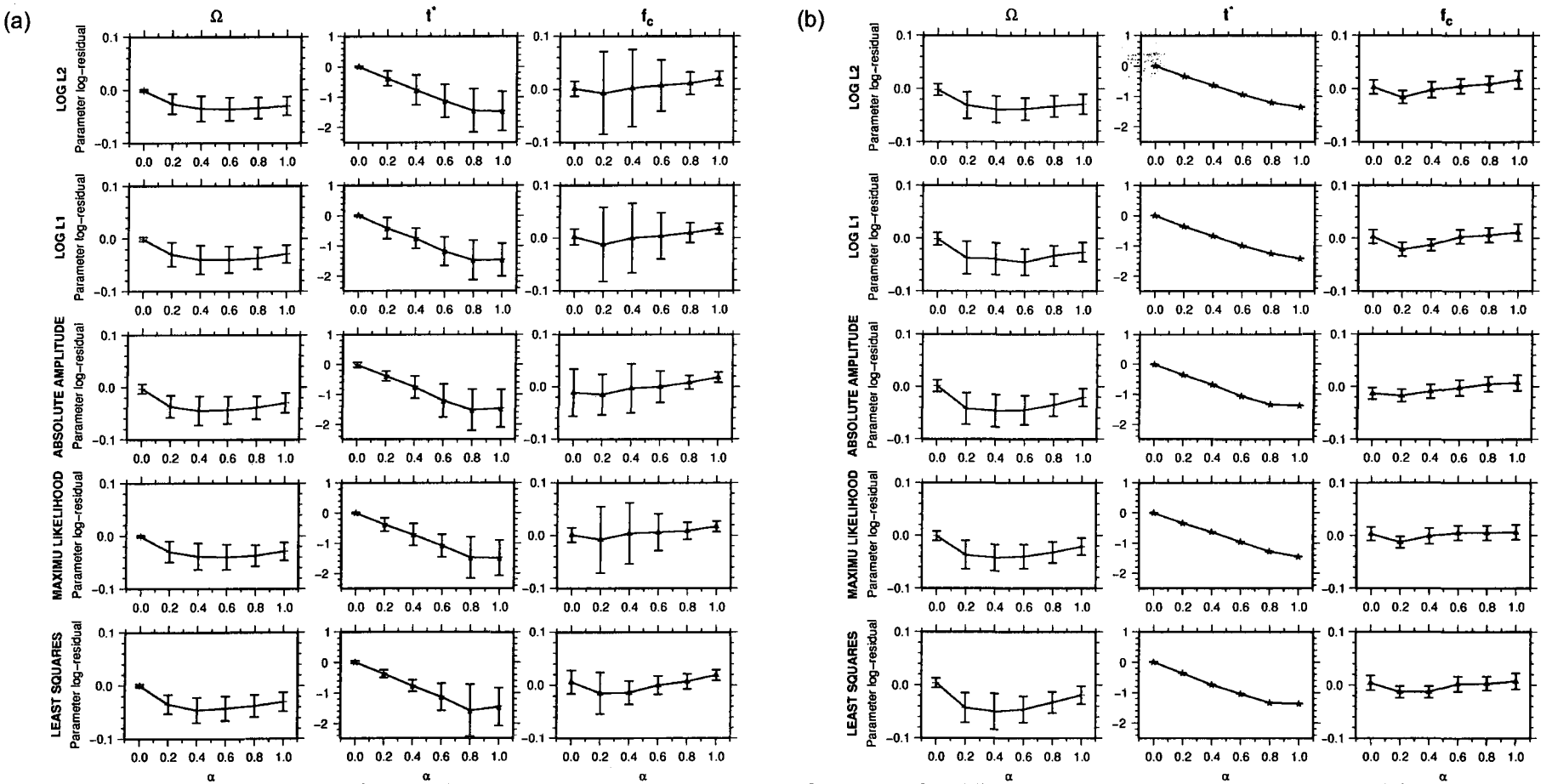

Figure 1.18: Parameter error for signal moment, Ω; t and source corner frequency, f_c, whilst varying the frequency dependence of Q, α, in the synthetic data. (a) before fixing Q, (b) after fixing Q. The inversion assumed α=0.0. Error bars are shown for one standard deviation.*

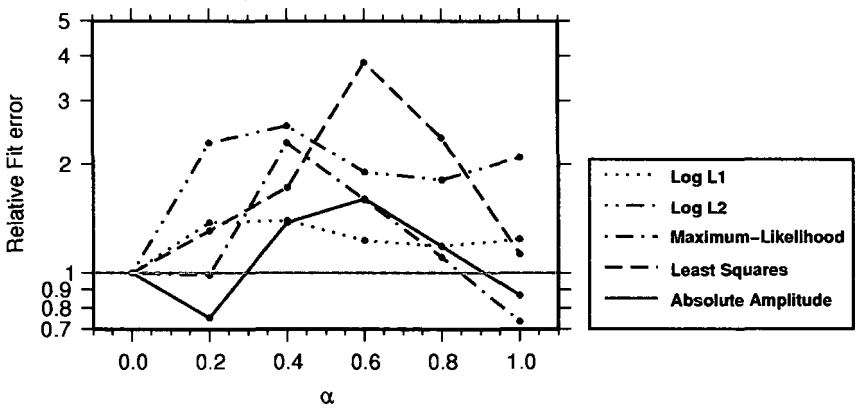

Figure 1.19: Plot of overall fit-residual against synthetic α when assuming α=0.0 for different minimisation functions.

In Figure 1.18 the parameter errors are shown for all different inversion types with each row representing a different minimisation function (as labelled). There is very little variation in the errors, dependent on this minimisation function, particularly after fixing the Q structure.

We can begin to understand the trade-off between α and the other parameters by considering that as α tends to unity, the degree to which the data can be modelled by a frequency dependent exponential function decreases:

$$\alpha \to 1.0 \Rightarrow e^{(-\pi \cdot f^{1-\alpha} \cdot t_{ij}^{*})} \to e^{(-\pi \cdot t_{ij}^{*})} \quad . \tag{1.55}$$

Such that when α equals one, the attenuation exponential becomes completely independent of frequency and is mapped into the signal moment parameter by the inversion. In the example above, where Q is modelled as frequency independent, the inclusion of α_{input}=1.0 should lead to no errors in t* , and the error in the signal moment to be:

$$\left.\begin{aligned} \xi_{\Omega_i} &= \frac{\Omega_i - \Omega_i \cdot e^{(-\pi \cdot t_{ij}^{*})}}{\Omega_i} \\ \xi_{\Omega_i} &= 1 - e^{(-\pi \cdot t_{ij}^{*})} \end{aligned}\right\} \alpha = 1.0 \ , \tag{1.56}$$

Such that the signal moment appears to be:

$$\Omega_i = \Omega_i (1 - e^{(-\pi \cdot t_{ij}^{*})}) \quad . \tag{1.57}$$

The corner frequency is therefore unaffected when α_{input}=1.0. In Figure 1.18 we can see some error in f_c remains at *α=1.0,* this is due to the grid-search resolution not resolving the

exact value of f_c. For instance, compare this with the error and scatter on f_c at α=0.0, where we expect zero error in all parameters (because we have a unique solution).

For $\alpha_{input}<1$ the effect of α is to cause a gradual change from being frequency dependent, $e^{(-\pi \cdot f \cdot t_{ij}^*)}$, to frequency independent, $e^{(-\pi \cdot t_{ij}^*)}$, (Equation 1.55). Up to α_{input}=0.4 this is increasingly accommodated by a lower signal moment and an increased scatter in the f_c error. For α_{input}>0.4 the signal moment increases toward the value given by Equation 1.57 at α_{input}=1.0, whilst the scatter in f_c error is reduced to zero (or as low as is limited by the grid-search).

The degree of apparent geometrical decay is affected by α_{input} (Figure 1.20). The trend of errors with α_{input} follows that of the errors in signal moments, with a peak at α_{input}=0.2 or α_{input}=0.3 depending on the minimisation function.

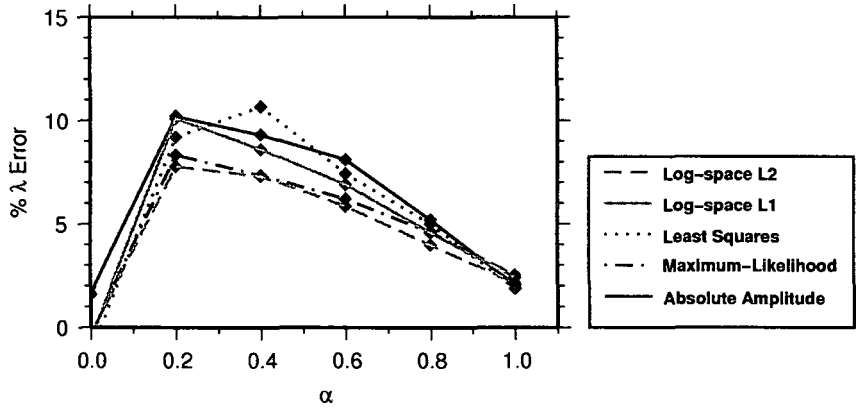

Figure 1.20: Error introduced into the apparent geometrical decay exponent due to the inclusion of frequency dependent Q, while modelling α=0.0.

This shows that the errors in signal moments are not uniform over all distances (in which case the degree of non-intrinsic attenuation would not change), rather that they are dependent on the source-receiver distance, and hence degree of attenuation. This observation supports Equation 1.57 which shows that in the case where α_{input}=1.0 the observed signal moment (assuming α_{mod}=0.0) is dependant on the degree of path attenuation, t*. As modelling a frequency dependent Q dataset with frequency independent Q (α=0.0) leads to the appearance of a higher rate of geometrical spreading. It is logical therefore, to assume that if the geometrical spreading function we use is too weak, that we will see high levels of frequency dependence of Q; even if Q is really independent of frequency. Caution must be used, that high levels of α, particularly when fixing the rate of non-intrinsic attenuation, are not results of mapping the geometrical spreading into α and, equally, vice-versa.

1.6.2.2 Quality Factor, Q

Figure 1.12 shows that within a certain frequency range it was possible to account for an incorrect value of Q_{mod} by altering the source spectrum shape, defined by its corner-frequency. This test aims to observe how the inversion process varies the other parameters to account for the incorrect value of Q in the inversion model. The influence of noise in the spectrum is again ignored, so that the explicit effects of varying Q can be analysed. Figure 1. 21 shows these changes for the different minimisation functions. As t* is proportional to the reciprocal of Q, the error induced on the t* parameter is proportional to the difference between the modelled and input Q. As the inversion model assumes that Q_{mod}=350; for Q_{input}>350 t* will be over-predicted and for Q_{nput}<350 t* will be under-predicted. Under-predicting t* is counter-acted in terms of minimising the data-model fit by reducing both the source corner frequency and the signal moment, and vice-versa for the over-prediction of t*. The exponential effect of Q can be seen in that for high values of Q the changes in parameter error are less significant. As such a doubling of Q from 500 to 1000 influences the other parameters only as much as changing Q from 200 to 400.

When significantly under-predicting t* (e.g., Q_{input} < 100; Q_{mod} = 350) and using the log-space inversions the signal moment is dramatically over-predicted. For instance, using the log-space L2 inversion the signal moment residual was 1.228 +/- 1.275 for Q_{input} = 100. In the case of the L2 minimisation the high frequency data contribute less to the overall fit as they are so small in amplitude. The model is therefore fit around the spectral peak, and is less affected by the strongly attenuated high frequency data. In the case of the log-space fit, the data are more-or-less equally weighted over the entire frequency range. To attempt to replicate the high attenuation the model uses a very high signal moment and a low corner frequency, allowing the correct modelling of the high frequency data at the cost of the low frequency data, where, in the log-space, we have competitively fewer data points.

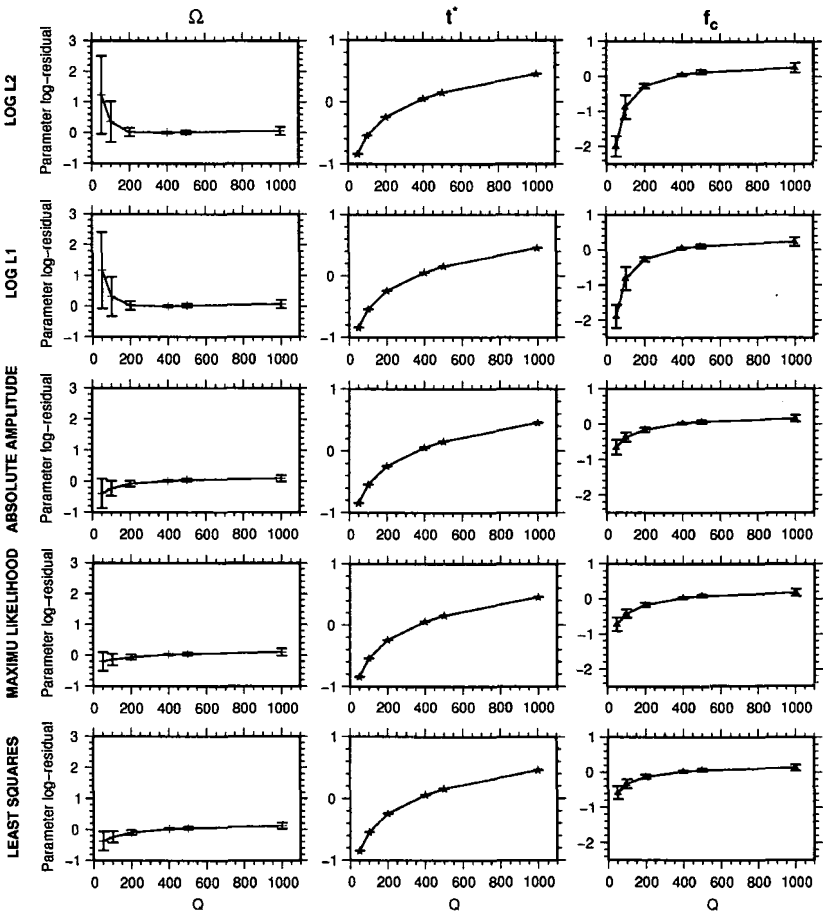

Figure 1.21: Parameter error for signal moment, Ω; t and source corner frequency, f_c, for different values of Q_{input} used in computing the synthetic data. The inversion assumed Q_{mod}=350. Error bars are shown for one standard deviation.*

1.6.2.3 Noise and Signal-to-Noise Ratio

Noise was applied to the model given by Equation 1.50. Values of p_1=0.1 and p_2=0.2 (equivalent to a mean noise level at 10% of the spectral maximum with 20% standard deviation) are broadly consistent with micro-earthquake recordings on short period (1s) seismometers. This was checked by looking at the noise and signal estimates of a British Geological Survey (BGS) database: signal-to-noise spectral levels were measured at the signal's peak spectral amplitude and were found to be:

$$\log(SNR) \approx \log(7) \pm 0.5 \tag{1.58}$$

for over 500 recordings. This is equivalent to a range between approximately 2 and 20 in the log-space. Figure 1.22 shows that as the noise level in the data, p_1, is increased we observe a decrease in t* and f_c. This is due to the flattening out of the velocity spectrum close to the noise level (e.g., Figure 1.14).

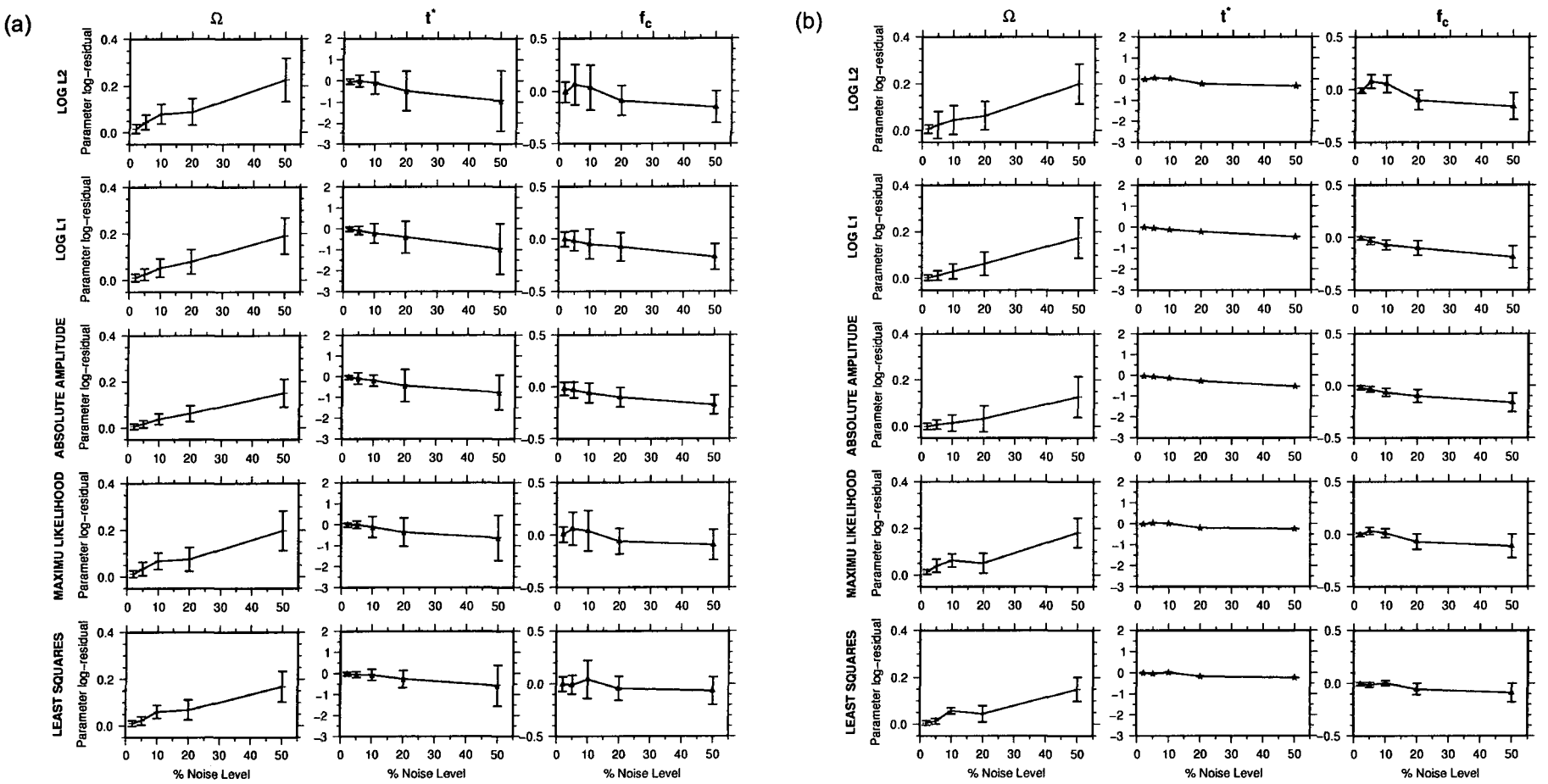

Figure 1.22: Parameter error for (left to right) signal moment, Ω; t and source corner frequency, f_c. (a) before fixing Q and (b) after fixing Q whilst varying the noise level (p_1x100%) of the synthetic data. Error bars are shown for one standard deviation.*

The effect of p_1 is the most significant on t*, with a decrease of around 50% due to p_1=0.5 (50% of the spectral maximum). In contrast the source corner frequency is reduced by around 20% at the same noise level. In addition to the affect on these parameters, which try to recover the shape of the noisy spectrum, the signal moment is increased by a factor approximately equal to the degree of noise being applied: at p_1=0.5 we see a 0.5 fractional increase in the signal moment. Contrary to this general trend, the L2, log-L2 and maximum-likelihood inversions produce a small initial increase in f_c for low noise levels between 0.02 > p_1 > 0.05.

In order to minimise the effect of the noise in the spectrum we impose a signal-to-noise ratio (SNR) cutoff. When we impose a SNR restriction we would assume that this would result in a reduction in the errors. However, there exists a balance between errors imposed due to the level of noise allowed in the data, and the spectrum frequency bandwidth. Too much noise and t* and f_c will be, on average, underestimated. Too little bandwidth and we increase the significance of the trade-offs between parameters. This is highly data dependent: for particularly noisy data, we may have to accept using lower than ideal SNR, so that a large enough frequency bandwidth is available. Imposing both high SNR and bandwidth restrictions will result in few data passing through to the inversion, resulting in large bias due to, for example, source receiver distribution or source mechanisms. The trade-off between data quantity and quality is exemplified by Figure 1.23. In this test the SNR was increased from 1 to 5 for whilst modelling the same data. This resulted in a change in average frequency bandwidth as shown in Figure 1.24. By increasing the SNR from 1 to 3 we observe a small decrease in the scatter of signal moment error. Negligible effects are present in t* and f_c errors in all inversions apart from the log-space L2; where a reduction in the amplitude of the mean error is achieved for both parameters. The effect of the bandwidth reduction is shown when increasing the SNR from 3 to 5. This results in an increase in the scatter of error on all of the parameters, in particular t* and f_c (e.g., Figure 1.23). The scatter on all of the errors is reduced after fixing the Q structure. However, the inversions using a SNR of 5 are still comparatively worse, in particular the log-space inversions; where the t* value is around 30-40% overestimated. This introduces error into the f_c parameter of around 60-80%.

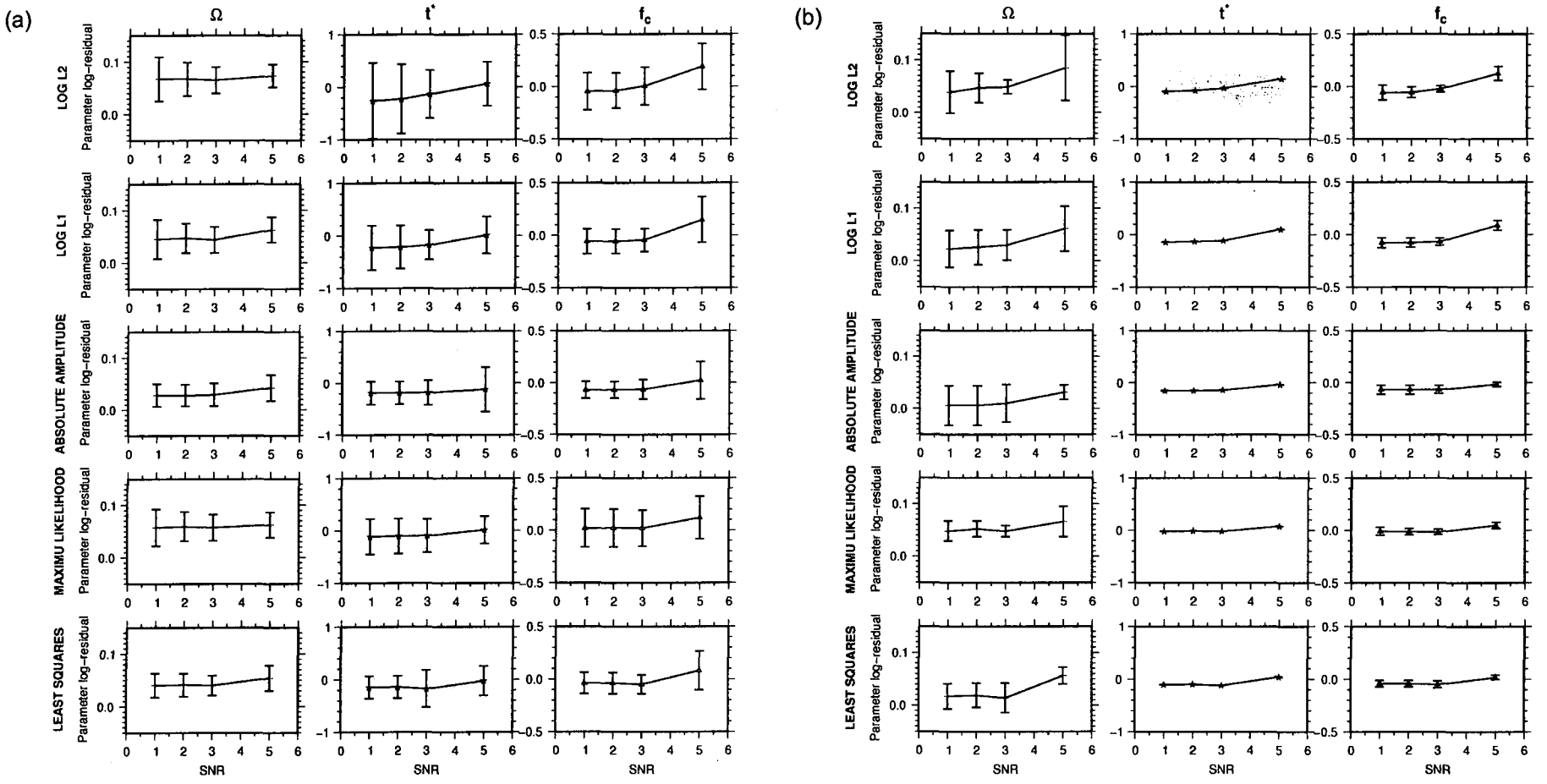

Figure 1.23: Parameter error for (left to right) signal moments, Ω; t and source corner frequency, f_c, (a) before fixing Q and (b) after fixing Q with a varying SNR imposed during data selection. Error bars are shown for one standard deviation.*

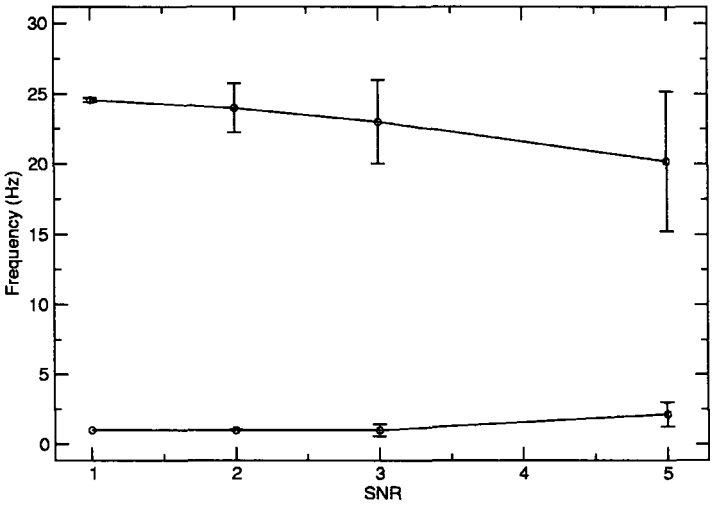

Figure 1.24: Change in the frequency bandwidth of the spectra as the SNR is increased (lower and upper bounds are plotted). Error bars show one standard deviation from the mean.

1.6.2.4 Source Spectrum Shape

We observe significant deviations from true values for all parameters as γ, the source spectrum falloff, deviates from the inversion model value of 2.0 (Figure 1.26). Two distinct trends can be observed: as γ decreases from 2.0 and as γ increases from 2.0. For both trends the use of $\gamma \neq 2.0$ in the synthetic data is mapped into errors in all parameters. However, the distribution of those errors in t^* and f_c is different in each case: for $\gamma > 2.0$ we see larger scatter in the errors in f_c compared to t^*, and vice versa for $\gamma < 2.0$. The Inclusion of γ adds a further frequency dependent component to the data. To some degree this component is taken up by the other modelled parameters. However, the ability to model the data with $\gamma \neq 2.0$ is poor. Although high errors are produced in all parameters, this still does not produce a good overall fit (Figure 1.25), unlike for $\alpha \neq 0.0$. We should therefore able to resolve this parameter using a grid search.

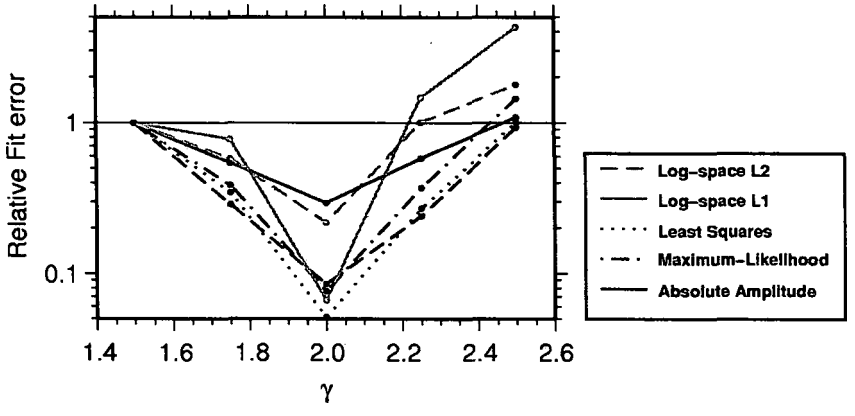

Figure 1.25: Plot of overall fit for various modelled source spectra, defined by the high frequency spectral decay, γ. For the synthetic source γ=2.0. The modelled γ was varied giving different degrees of residual fit.

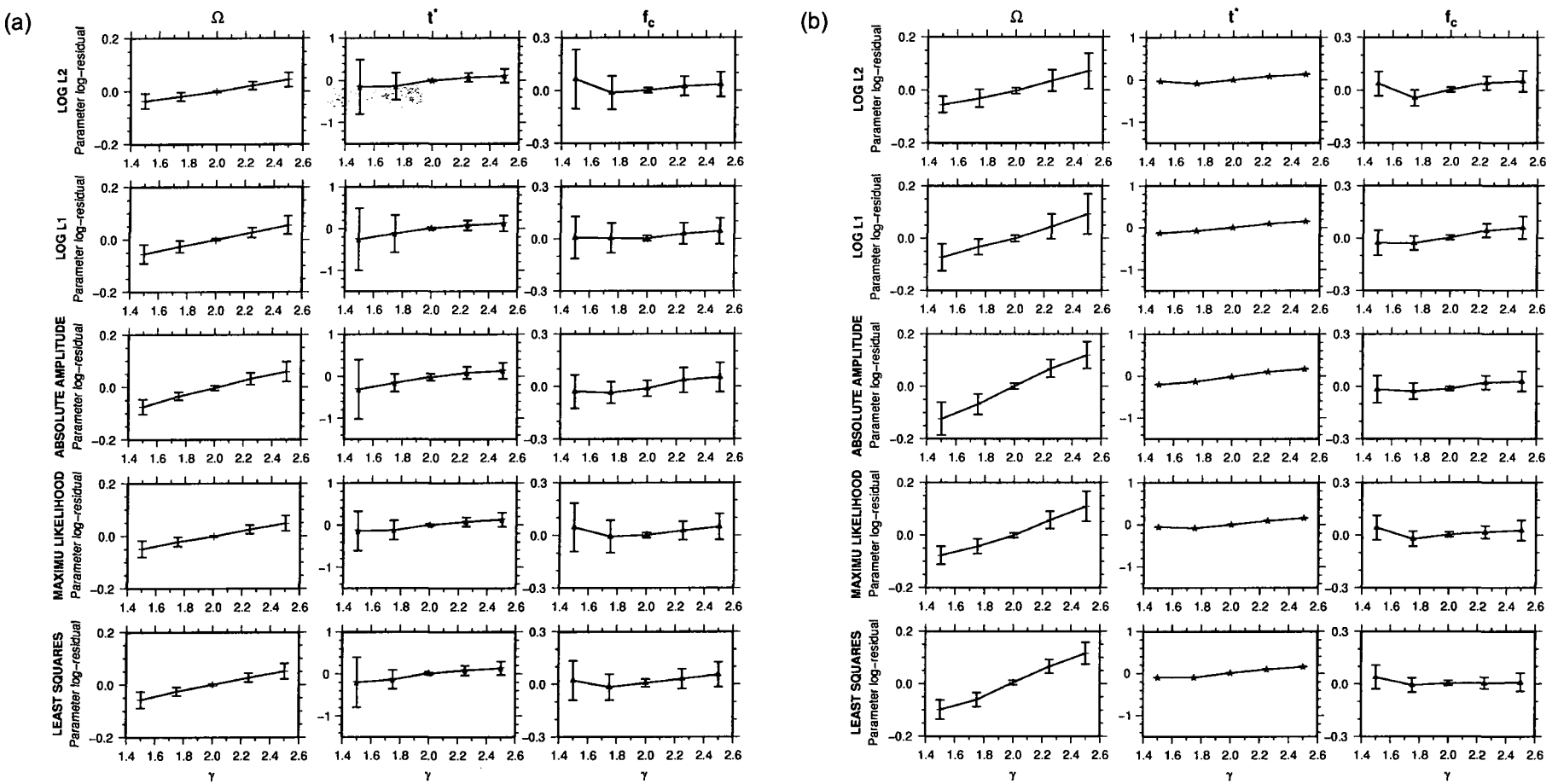

Figure 1.26: Parameter error for signal moments, Ω; t and source corner frequency, f_c, (a) before fixing Q and (b) after fixing Q whilst varying the source spectrum falloff, γ, of the synthetic data. Error bars are shown for one standard deviation.*

1.6.3 Summary

In this chapter a model of earthquake energy propagation in the far-field was introduced. The methodology used to deconvolve the parameters of this model from recordings of earthquakes was also detailed. By analysing both stages of this inversion (before and after applying constraint to the Q structure) we tested the success and limitations of the method. We observed that by applying the fixed Q-structure constraint the scatter of the errors in f_c are significantly reduced. In particular when applying noise to the synthetic spectra (e.g., Figure 1.22). For instance, at a noise level similar to that found in the dataset of the British Geological Survey (Values of p_1=0.1 and p_2=0.2: equivalent to a mean noise level at 10% of the spectral maximum with 20% standard deviation) the log-residual of f_c was reduced from 0.037±0.213 to 0.056±0.083 for the Log L2 minimisation. However, we also found that any bias in the original t* estimates is propagated through to the Q structure (e.g., Figure 1.21). For instance, we observed that by underestimating the initial t* value by a certain percentage, that this results in approximately the same error being propagated into f_c. This suggests that although the method acts to form a robust solution, we must be careful to initially obtain an unbiased estimate of the t* values. We saw that the frequency dependence of Q was not well resolved, even with noise-free data. This was exemplified by the fact that for values of $\alpha \neq 0.0$ the relative increase in the residual misfit for the Log L2 fit was only a factor of 2 (Figure 1.19). Considering that this was a noise-free test this is a low value: ideally we wish to obtain a steeply increasing gradient in the relative fit away from the correct model value in order to be able to resolve the parameter (e.g., Figure 1.25). Additionally there was a trade off between the frequency dependence of Q and the apparent geometrical decay due to the distance dependence of the error in the signal moments (e.g., Equation 1. 56). This suggests that the relatively common assumption of a 1/R geometrical decay could influence the apparent frequency dependence of Q if, in reality, this assumption is not correct. The source spectrum shape, on the other hand, was much better resolved (Figure 1. 25). In this case the inversion parameters were less able to accommodate the incorrect source spectrum. For instance, a ±0.5 change in the value of γ, the high frequency decay of the source, led to a factor of 10 increase in the residual fit. The balance between bandwidth and the signal-to noise ratio is important to consider. It was shown that higher signal-to-noise ratios do not necessarily lead to better solutions due to the loss of bandwidth in the signal (Figure 1.23).

There are limitations with these tests. The results are likely to be dependent on the dataset used, although common trends will remain the exact performance will depend on factors such as the number of observations per event, the range of magnitudes and hypocentral distances. It is therefore important to perform tests on the dataset being studied. Additionally, these tests do not investigate the validity of the point source model assumption (Brune, 1970, 1971; Boore 2003), or the possibility of strong site effects in the record. This will be addressed in the following Chapter.

Chapter 2

Testing of method using synthetic seismograms

2.1 Introduction

In the following chapter the model and technique from Chapter 1 is tested through its application to synthetic seismograms. This builds upon the basic synthetic testing in Chapter 1 by beginning simulation with the time-series seismogram rather than directly modelling the Fourier spectra to produce a synthetic dataset. The motivation of this chapter is twofold. The first is to analyse how the inversion performs when we know the source mechanism including the true stress drop and seismic moment associated with this source. In this way we will be able to show how the results of the methodology presented in Chapter 1 are associated to the rupture mechanics of an earthquake. Additionally this chapter aims to test for self-consistency between the method in Chapter 1 and the method of Boore (2003), used to simulate measures of strong ground motion, such as peak velocity or acceleration. This is important, because the method of Boore (2003) requires the results obtained using the method in Chapter 1. For instance we need to be sure that the definition of the Fourier spectrum is the same (e.g., one sided or two sided spectrum), whether the geometrical decay function is compatible and how the stress drop is defined. Making the assumption that the two models are compatible without such testing can lead to erroneous predictions of ground-shaking. In addition to this, using the method of Boore (2003) we are able to simulate the site response (e.g., Steidl et al., 1996), an element to which our method may be susceptible.

In the first section of this chapter synthetic seismograms are produced using a finite-difference method to solve the system of first order wave equations over a staggered 3D grid (e.g., Madariaga, 1998). This means that a range of fault types and sizes can be simulated to give realistic seismograms recorded at a distance from the fault. The programme WAVE (Cundall, 1992; Hildyard, 1995) is used to produce these synthetic seismograms. This facilitates the comparison between the rupture mechanics and the attenuation parameters derived using the method in Chapter 1.

In a second section of the chapter synthetic Fourier spectra are produced using the stochastic method of Boore (2003). This method utilises the random nature of the observed ground-motion. A theoretical spectrum is produced based on the instantaneous slip model of Brune (1970, 1971) along with propagation effects: geometrical spreading and attenuation. This is then multiplied with the FFT of a random noise envelope before returning to the time domain. The programme SMSIM (Boore, online) is used for this simulation.

2.2 Synthetic Data from finite-difference modelling (WAVE).

This method provides realistic seismograms simulated using 3D fault planes. The simulation takes into account the complex interaction of the wave-field with the fault edges and the surface interface. To do this it solves the system of first order wave equations over a

staggered 3D grid. Faults are constructed from smaller sub-faults, each of which ruptures independently. This allows for realistic faults and rupture propagation effects to be considered. Figure 2.1 shows the location of the sensors relative to the fault. Table 2.1 summarises the five events simulated using the finite-difference method. The earthquakes ranged in size from just under M_w=3.5 to M_w=6.4. Two sensors were used, and the ground motion on three components (vertical, north-south and east-west) were simulated.

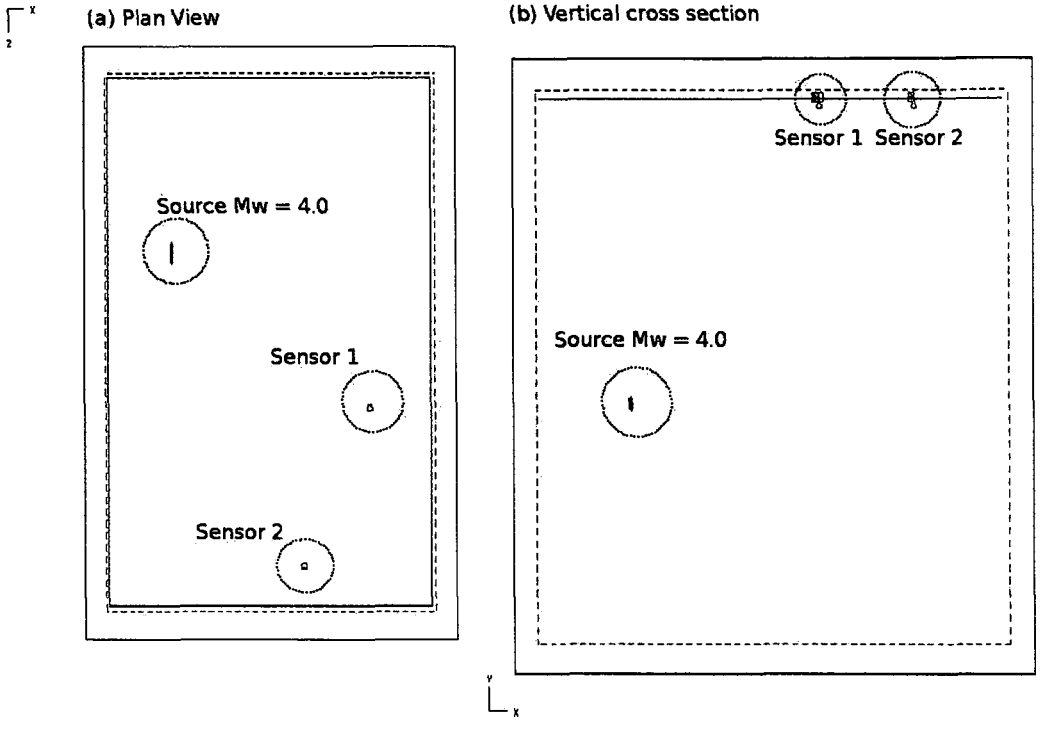

Figure 2.1: (a) Plan view and (b) cross section of the source (M_W = 3.95) and sensor locations for the case of the M_w = 4.0 event. The fault dimensions are (a) 2km by (b) 1km in this case.

No.	Mw	Fault Width, W (km)	Fault Length, L (km)	Horizontal Rupture Time (s)	Vertical Rupture Time (s)	Fault Depth (km)	Hypocentral Dist. (km) Sensor 1	Sensor 2
1	3.45	1.2	0.6	0.20	0.10	30	46	46
2	3.95	2.0	1.0	0.33	0.17	30	46	46
3	4.98	6.4	3.2	1.07	0.53	35	49	50
4	5.94	19.2	9.6	3.20	1.60	25	43	44
5	6.41	32.0	16.0	5.33	2.67	25	43	44

Table 2.1: Summary of events simulated using the finite-difference method showing magnitude, fault geometry and location. Fault length and width are the horizontal and vertical extent respectively.

The parameters used in the simulation programme are shown in Table 2.2. In all cases, the rise-time for each sub-fault is 0.2s.

Element size	0.2 km
Time step	0.0136 s
P wave velocity	6.61 km/s
S wave velocity	3.865 km/s
Rupture velocity	3.0 km/s
Rise time	0.2 s

Table 2.2: Summary of parameters used for simulating events listed in Table 2.1 using the programme WAVE. The grid-mesh is defined by the element size and the simulation is performed at intervals defined by the time-step. The rupture propagates at the given rupture velocity with each sub-fault in the grid taking the time given by 'rise time' to complete its slip. Larger earthquakes are achieved by increasing the number of elements used.

The simulation was performed by M. Hildyard using the programme WAVE (Cundall, 1992; Hildyard, 1995). The simulated seismograms provided by Hildyard were windowed by taking the data comprising the 5 to 95% energy integral of an 8 second window centred on the peak velocity. The use of an 8 second window, refined to include only the 5 to 95% energy integral contribution was found to be suitable to encompass the direct S wave arrival(s) (Figure 2.2), whilst avoiding the inclusion of the P and S wave codas. The energy integral method (e.g., Raoof et al, 1999) allows the distance and magnitude dependence of the duration of shaking to be taken into consideration. The S wave is chosen due to it comprising of the peak ground-shaking and therefore being of significance in hazard assessment. The resulting waveforms were then processed using a multi-taper method (Lees and Park, 1995; Park et al., 1987) using 3π-prolate tapers. A FFT (Press et al, 1997) was then applied to the tapered, windowed time series giving the velocity spectrum of each component. The two horizontal components are combined through a vector sum of the individual N-S and E-W components.

We used only the frequency band up to 1.6Hz as this is below the maximum frequency at which the method was expected to be valid (Hildyard, pers. comm.). This maximum frequency is limited by the simulation method. Roughly speaking the minimum valid wavelength is given by around five to ten times the grid spacing, in this case the minimum wavelength is therefore 2km. For the S-wave this equates to a maximum frequency of around 2.0 Hz.

An example of the simulated seismograms can be seen in Figure 2.2. In this figure we can actually observe that after the theoretical limiting frequency of ~2Hz the amplitude decay with increasing frequency slows due to invalid high-frequency data being added to the seismogram. For the larger events this can be observed occurring at less than 2Hz, more like the 1.6Hz we use as a limit in the inversion modelling. A further limitation of the data from this simulation is that only two sensors were available, both of which lie at similar azimuths from the source.

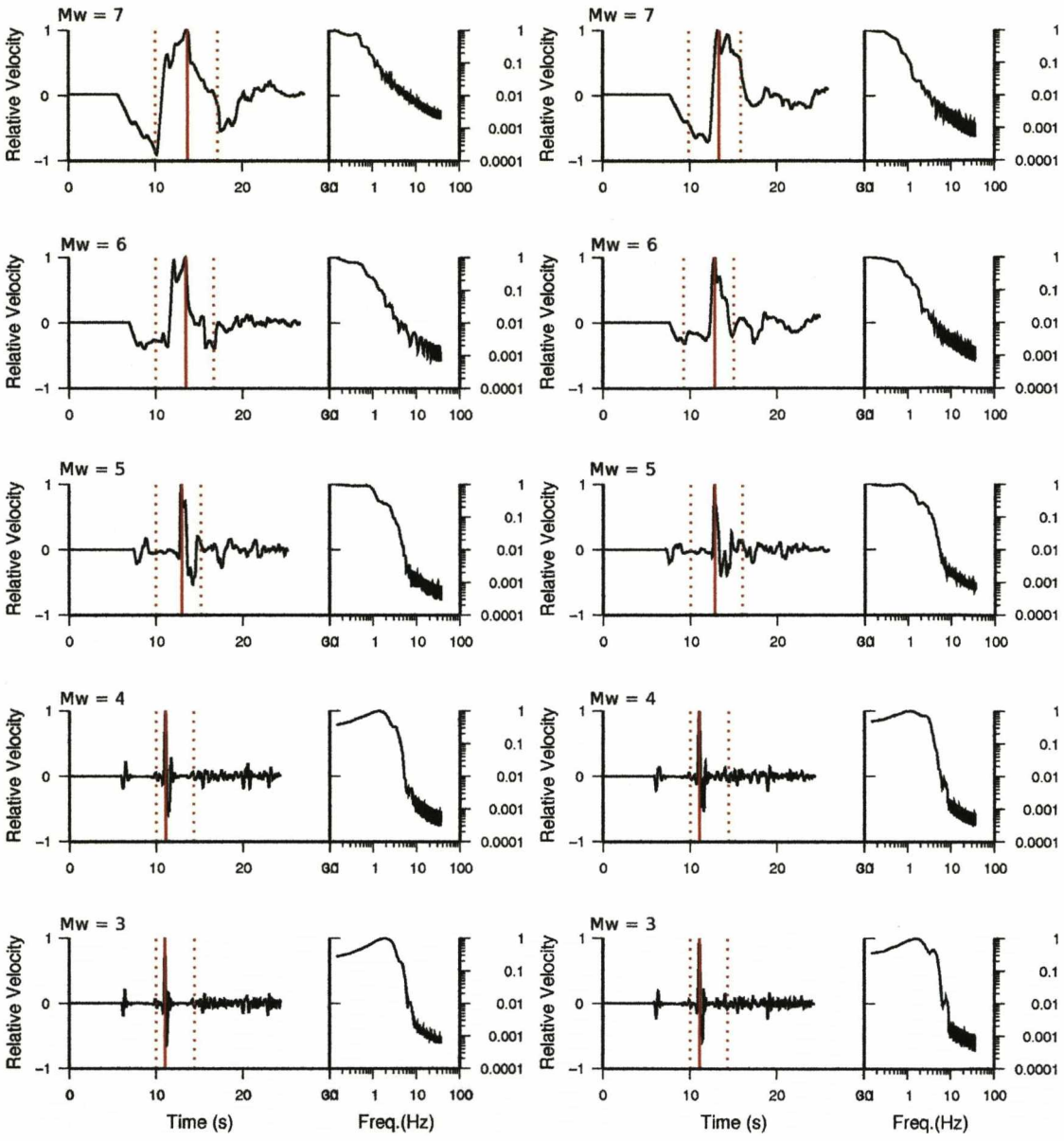

Figure 2.2: Plot of the vertical component synthetic seismogram and the multi-taper FFT at sensor 1 and 2 for each of the 5 events in Table 2.1. The solid red lines indicate the maximum amplitude, the dashed red lines indicate the beginning and end of the signal windows used to compute the spectra. Earthquake 1 is at the bottom, with sensor 1 on the left and 2 on the right, through to earthquake 5 at the top. The plots show relative velocity: in the case of the time series this is relative to the peak ground velocity and in the case of the frequency spectra relative to the peak Fourier velocity.

2.2.1 Inversion Method

We follow a slightly adapted version of the method described in Chapter 1 for the inversion of the velocity spectra. To briefly summarise, in this method the spectral velocity at any given hypocentral distance and frequency is given by:

$$\Omega_{ij}(f,R)=\frac{2\pi\Psi_i\cdot A_j\cdot a_j(f)}{R\cdot\left(1+\left(\frac{f}{f_c}\right)^{n\gamma}\right)^{1/n}}e^{-\pi f\frac{T_{ij}}{Q(f)}+\kappa_j} \quad . \tag{2.1}$$

We set $\gamma=2.0$ and n=1.0, which is equivalent to the Brune ω^2 source model (Brune 1970, 1971). The frequency independent terms are collectively termed the signal moment, $\hat{\Omega}_{ij}(r)$, given by:

$$\hat{\Omega}_{ij}(R)=\frac{\Psi_i\cdot A_j}{R} \quad . \tag{2.2}$$

In Chapter 1 we detailed how we invert for the parameters of this model by first constructing a Q model using estimates of the $t^*_{ij(comb)}$ parameter, given by:

$$\frac{T_{ij}}{Q(f)}+\kappa_j=t^*_{ij(comb)}=\int_{raypath}\frac{1}{Q(x,y,z)\cdot v(x,y,z)}ds+\kappa_j \quad . \tag{2.3}$$

where $t^*_{ij(comb)}$ is a measure of the total path attenuation, v(x,y,z) defines a velocity model, Q(x,y,z) defines the Q model and κ_j is a measure of attenuation attributed only to the recording site. This model is an important part of the stabilisation of the inversion. The initial $t^*_{ij(comb)}$ estimates are obtained from a combined grid-search through f_c, the source corner frequency, and a Powell's search algorithm for the signal moment (Equation 2.2) and the $t^*_{ij(comb)}$ parameter. In this case, however, our synthetic data are only valid up to around 2Hz, being limited by the method of simulation. At these low frequencies the effect of Q is not even apparent for realistic values of Q, hence we cannot resolve it in this dataset. As a result we chose to skip the first part of the method in Chapter 1 and make use the fact that we know the correct value of $t^*_{ij(comb)}$. In this simulation $t^*_{ij(comb)}=0$, which implies no intrinsic attenuation: the simulation is performed using an elastic medium. The resultant method applied to this dataset is therefore simply a grid-search (with 5% resolution in this case) through the source corner frequency , f_c, and a direct regression for the frequency independent amplitude parameter, the signal moment ($\hat{\Omega}_{ij}(r)$). In Chapter 1 this spectral inversion is followed by a SVD decomposition of Equation 2.2 in the log-space. However, since we have very few data points in this case, all at approximately the same distance from the source, we provide the rate of geometrical decay as equal to 1/R. Additionally we fix A_j=1, such that there is no site amplification in this case. This simply leaves a conversion from the signal moment to the seismic moment. This conversion is given by:

$$M_{0i} = \frac{\Psi_i \cdot v^3 \cdot \rho \cdot r_0 \cdot 4\pi}{\xi F \Theta_{\lambda\phi}}, \tag{2.4}$$

(Brune, 1970). Where M_{0i} is the seismic moment (in SI units), ψ_i is the long-period plateau value at source i, and $\Theta_{\lambda\Phi}$ is the average radiation pattern ($\Theta_{\lambda\Phi}$ = 0.55 for S waves (Aki and Richards, 1980; Boore and Boatwright, 1984)). v is the S-wave velocity at the source (we assume on average v=3.5km/s), F is the free surface amplification (F = 2.0 for normally incident SH waves and a good approximation for SV) and ρ is the average crustal density (ρ= 2800kgm^{-3}). ξ is a factor to account for the partitioning of energy onto the vertical plane (e.g., Newmark and Hall, 1982), for one component of the horizontal plane $\xi \approx \sqrt{2}$ and for the vector sum of the two horizontal components $\xi \approx 1$. r_0 is the rupture radius which can be estimated by assuming a circular rupture using:

$$r_0 = \frac{v_s k}{f_c}, \tag{2.5}$$

(Brune 1970). For shear waves, v_s is the shear wave velocity at the source, and k is 0.37. The moment magnitude can then be calculated using:

$$M_w = \frac{2}{3}\log M_0 - 6.03, \tag{2.6}$$

along with the stress drop:

$$\Delta\sigma = \frac{7}{16} \cdot \frac{M_0}{r_0^3}, \tag{2.7}$$

(Eshelby, 1957), assuming a circular rupture plane. Whilst this is not actually correct in this case it is how we approach the problem without prior knowledge of the fault geometry.

2.2.2 Results

The inverted magnitudes and stress drop are shown in Table 2.3 for the vector sum of the two horizontal components and the vertical component for data between 0 and 1.6Hz. The stress drops of the events are estimated using the following equation:

$$\Delta\sigma = \frac{2M_0}{\pi W^2 L}, \tag{2.8}$$

where L and W are the fault length (horizontal) and width (vertical) as given in Table 2.3. This assumes a rectangular rupture plane with a strike-slip movement. Alternatively for a dip-slip event:

$$\Delta\sigma = \frac{8M_0}{3\pi W^2 L} \quad , \tag{2.9}$$

Simulation						Inversion			
						Horizontal Component		Vertical Component	
Mw	Mo (N.m)	Length (km)	Width (km)	Dip-Slip σ (Mpa)	Strike-Slip σ (Mpa)	Mw	σ (Mpa)	Mw	σ (Mpa)
3.45	1.88E+014	1.2	0.6	0.28	0.37	3.5	0.12	3.59	0.09
3.95	1.06E+015	2.0	1.0	0.34	0.45	4.01	0.33	4.11	0.25
4.98	3.72E+016	6.4	3.2	0.36	0.48	5.19	0.23	5.16	0.32
5.94	1.02E+018	19.2	9.6	0.37	0.49	6.01	0.23	6.05	0.34
6.41	5.19E+018	32.0	16.0	0.4	0.54	6.4	0.28	6.48	0.26

Table 2.3: Fault geometry, magnitude and theoretical stress drop assuming a strike-slip or dip-slip rectangular fault, along with the inversion results from the vertical and horizontal components using data from 0 to 1.6Hz.

After applying the method described above to the data up to 1.6Hz we find that the inverted moment magnitudes are all within 0.2 of the correct value (Figure 2.3), regardless of the component used. The 4.98Mw event is the worst estimate, at 5.21 and 5.16 for the horizontal and vertical planes respectively, however this is still only 5% in error. The stress drop used in the simulation was 1MPa. Both equations 2.8 or 2.9 underestimate the stress drop by a factor of 2 to 3. The inverted stress drop, or more specifically the Brune stress parameter (Figure 2.4) was a factor of 3 to 4 lower than the input stress drop. However, the inverted stress drops are relatively constant apart from the M_w = 3.5 event. This is due to the corner frequency of the M_w = 3.5 event being 1.2Hz, close to the maximum of the frequency bandwidth available. There is minimal difference between the magnitudes computed from the horizontal and vertical data (Figure 2.3), suggesting that the use of the parameter ξ in Equation 2.4, which allows continuity between the different components of recording is correct. The radiation pattern coefficient was not known for this simulation, but from the results it seems that the average value given by Boore and Boatwright (1984) was adequate.

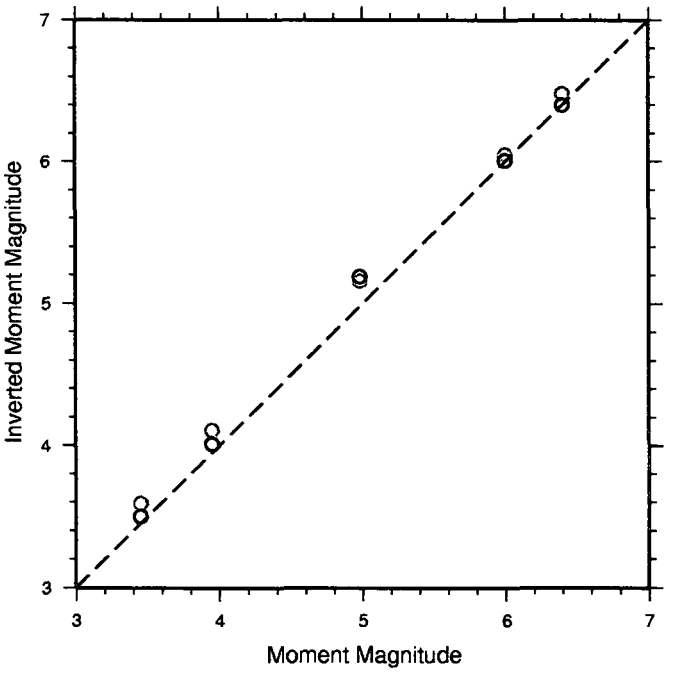

Figure 2.3: Plot of inverted moment magnitudes against actual moment magnitudes from the vertical (grey circles) and horizontal data (black circles). The dashed line is the 1:1 relationship.

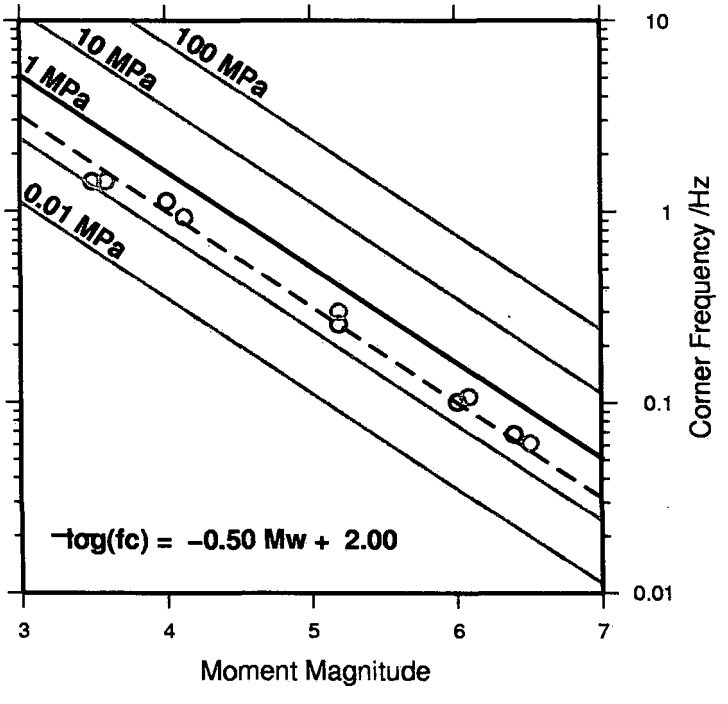

Figure 2.4: Plot of inverted stress drop against theoretical stress drop from the vertical (grey circles) and horizontal data (black circles) . The grey lines indicate constant stress drop and the dashed line indicates the mean constant stress drop for the horizontal data. The solid dark line indicates the true stress drop at 1MPa.

The residuals from the inversion of the horizontal data are shown in Figure 2.5. There is no dependence of the residuals on magnitude or distance (although the distance range is relatively short). The median residual is independent of frequency in the range 0 to 1.6Hz, but as the frequency approaches 1.6Hz the scatter begins to increase, indicating that this model is not as appropriate to the higher frequencies (i.e., greater than 1.6Hz).

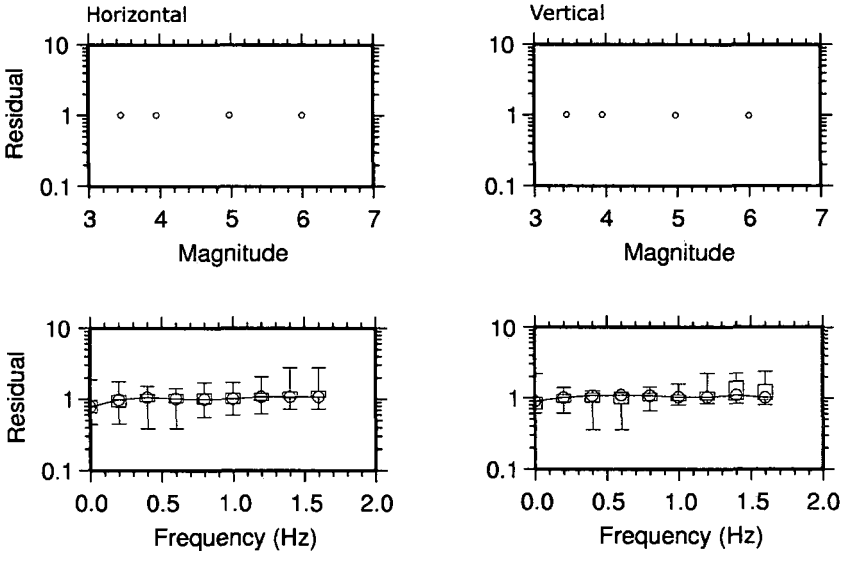

Figure 2.5: Residuals of the spectral inversion plotted against frequency and magnitude for the horizontal data (left) and the vertical data (right).

2.3 Fourier spectra simulated using the stochastic method.

It is also possible to simulate realistic Fourier spectra through the multiplication of a theoretical spectrum (e.g., Equation 2.1) with the an enveloped random noise time-series. This is the principal behind the stochastic method (e.g., Boore, 2003) which is commonly used to estimate peak ground-motions in regions of weak to moderate seismicity. The stochastic method of Boore (2003) simplifies the wave-propagation by assuming a point source fault (e.g., Brune 1970, 1971), using an average radiation coefficient to radiation effects (e.g., Boore and Boatwright,, 1984), and a factor of 2 to account for free-surface amplification.

2.3.1 Data

We compute 13 synthetic seismograms recorded at distances ranging from 10 to 200km for six magnitudes: 3 to 7 giving a total of 65 seismograms. The theoretical spectrum is given by:

$$\Omega_{ij}(f,r)=\frac{\xi F \theta_{\lambda\phi} M_{0i}\cdot a_j(f)}{4\pi R v^3 \rho\cdot\left(1+\left(\frac{f}{f_c}\right)^{n\gamma}\right)^{1/n}}\cdot exp\left(-\pi\cdot f\cdot\left(\frac{T_{ij}}{Q(f)}+\kappa_j\right)\right) \quad (2.10)$$

with the free surface amplification F=2, $\Theta_{\lambda\phi}$ =0.55 is the average radiation pattern. v is the S-wave velocity at the source (v =3500m/s) and ρ is the average crustal density (ρ = 2800kgm^{-3}). R is the hypocentral distance, and f_c is the source corner frequency. M_0 is the seismic moment (in SI units), given by:

$$\log(M_0)=1.5M_W+9.1 \quad (2.11)$$

ξ is set to 1 so we simulate all the energy on the observed plane. T_{ij} is the travel time of the S wave, Q(f)=1000 is the quality factor and κ_j =0 is the site attenuation term often termed kappa (e.g., Anderson and Hough, 1984). Different functions are used for $a_j(f)$. In the first case $a_j(f)$=1: this model is termed M1. Later sections deal with changing this function. Finally the source corner frequency is given by:

$$f_c=0.4906\beta\left(\frac{\Delta\sigma}{M_0}\right)^{\frac{1}{3}} \quad , \quad (2.12)$$

in SI units, where the stress drop, $\Delta\sigma$=1MPa.

As the forward model (Equation 2.10) is very similar to the inverse model we are using (i.e., a combination of Equations 2.1, 2.2, and 2.4), we should be able to directly recover all the individual parameters. It also required that the parameters that we obtain from the inverse problem are contiguous with those used in the stochastic model, as we later go on to use the simulation technique to synthesise large magnitude earthquakes. Examples of the data simulated using this method are shown in Figure 2.6. The duration of shaking using the stochastic simulation is determined by the distance from the source as well as the magnitude of the earthquake: for larger magnitudes and distances the duration of shaking is increased. This is the same as is observed for real earthquakes. It is of course critical that the S wave energy is encompassed by the analysis window, otherwise the computed moment magnitudes will be underestimated. It is important to consider the duration of shaking for each dataset, as this is likely to be regionally variable (e.g., Boore 2003). Due to a somewhat sparser distribution of the S wave energy over time in this stochastic simulation (compared to the finite difference modelling and empirical data used elsewhere in this thesis) it was found that longer analysis windows were required in order to contain the complete S wave energy.

Therefore rather than an initial 8 second window the whole trace is initially windowed. The windowed data are then refined to include 5 to 95% of the energy integral, as with the previous data.

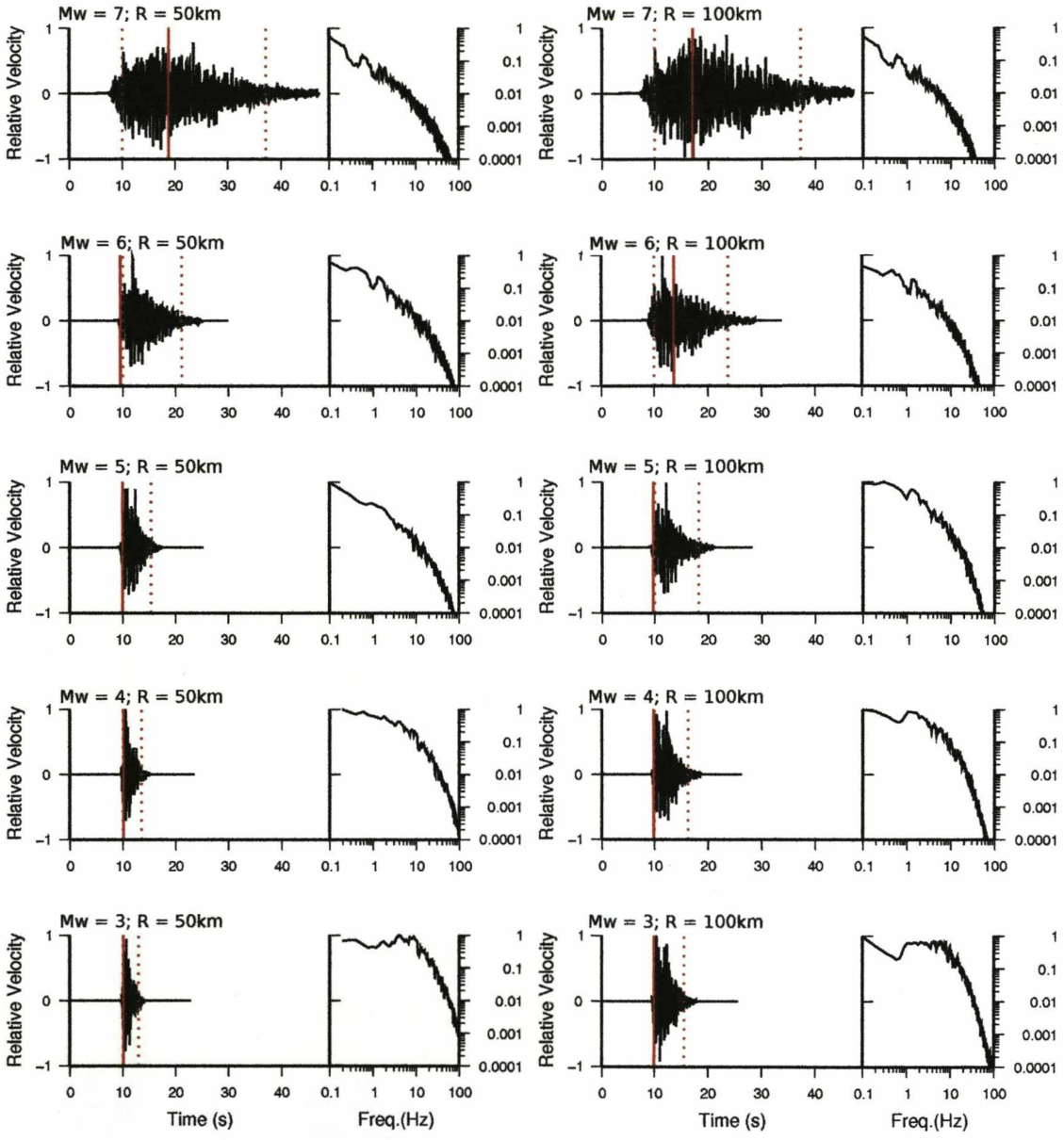

Figure 2.6: Simulated seismograms from earthquakes of M_W = 3 (bottom row) to 7 (top row) at 50km (left) and 100km (right) using the stochastic method. The solid red lines indicate the maximum amplitude, the dashed red lines indicate the beginning and end of the signal windows used to compute the spectra. The plots show relative velocity: in the case of the time series this is relative to the peak ground velocity and in the case of the frequency spectra relative to the peak Fourier velocity.

2.3.2 Forward Modelling

We initially test our model and technique are fully compatible with the programme SMSIM (Boore, online) used to simulate the seismograms. To do this we forward model the spectrum

using the same values in our model as for the simulation, as detailed above. We then compare this model against the Fourier spectra of the simulated seismograms. Figure 2.7 shows the residuals computed from the forward modelling of Equations 2.1, 2.2, and 2.4 with a 1MPa stress drop; Q(f) = 1000; κ_j=0; A_j=1 and $a_j(f)$=1. This model is termed M1 in the text. The residuals show that this model is perfectly satisfactory for modelling the simulated data. This shows that the methods are compatible, such that the results of the inversion of detailed in Chapter 1 can be used in the programme SMSIM to simulate large earthquakes.

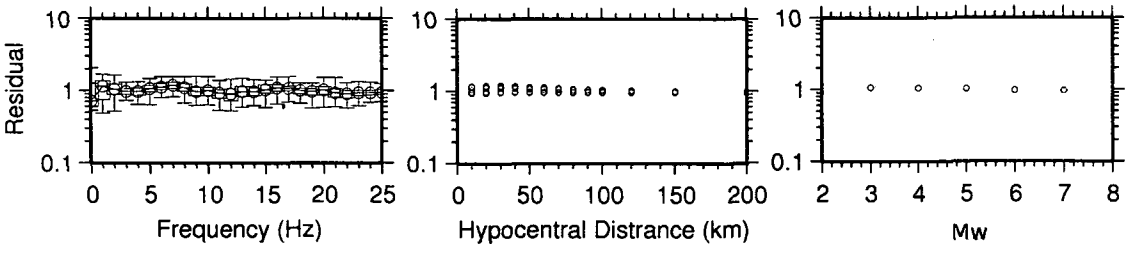

Figure 2.7: Residuals from the forward modelling of spectra (with M1) using the methodology in Chapter 1 relative to the synthetic seismograms computed using the stochastic method. For the plot against frequency the residuals are shown with the 63% and 95% confidence intervals (the box and whisker respectively).

2.3.3 Constructing the Q model

The next stage is to follow the method in Chapter 1, as summarised earlier in section 2.2.1, in which the parameters in Equation 2.1 are inverted for. This should allow us to test for any dependencies or limitations in the method. The data produced using the stochastic method are valid throughout the frequency range for which the theoretical spectrum is valid (in this case we use the frequency band from 0 to 25Hz). As such we do not have the problem as with the previous dataset (the finite difference model) and can more fully follow the method in Chapter 1. We first invert for estimates of the $t^*_{ij(comb)}$ parameter in order to reconstruct the Q model. In this case the input model was a homogeneous Q. Due to the small dataset and limited range of source depths and locations we also only use a homogeneous model in this inversion. Using the homogeneous half-space Q model we obtain a value of Q of 960 (to two significant figures) ±15 depending on the starting value of Q. This is very close to the input value of 1000. The station specific attenuation (κ_j) parameters can be negative or positive (as it is defined relative to the Q model) so there was no simplification involved in setting κ_j=0 in the synthetic data. In Table 2.4 we observe a small trend of increasing κ_j for sensors at greater distances from the source. This is also seen in the initial estimates of $t^*_{ij(comb)}$, with higher than expected values close to the source. This tendency to overestimate $t^*_{ij(comb)}$ near to the source is because the effect of attenuation is less obvious the nearer we are to the source and as such becomes difficult to resolve – the tendency highlighted here is that the attenuation in this case is overestimated, and it is likely that this is counterbalanced by an increase in the source corner frequency and reduction in the signal moment parameters (as shown in Chapter 1). The error on κ_j increases for stations further away from the source:

from ±0.002 for the station 10km from the source to ±0.043 for the station 200km from the source.

Sensor	Hypocentral Distance (km)	κ (s)	±σ	Resolution
1	10	0.004	0.002	0.98
2	20	0.003	0.004	0.98
3	30	0.002	0.006	0.98
4	40	0.001	0.009	0.97
5	50	0.001	0.011	0.96
6	60	0.001	0.013	0.95
7	70	0.000	0.015	0.94
8	80	0.000	0.017	0.93
9	90	0.000	0.019	0.92
10	100	0.000	0.021	0.9
11	120	0.000	0.026	0.86
12	150	-0.001	0.032	0.79
13	200	-0.001	0.043	0.64

Table 2.4: Recovered station attenuation parameters (κ_i) from the inversion of M1. The distance from the source to the sensor is shown along with the standard error and the parameter resolution value. The input κ_i=0.000.

The trend in the error suggests that the original estimates of $t^*_{ij(comb)}$ are consistently overestimated at short hypocentral distances, and at greater distances we observe a larger variability in our ability to resolve $t^*_{ij(comb)}$. However, on average the value of $t^*_{ij(comb)}$ is obtained on average to within 0.001 for all but the sensors in the first 30km.

2.3.4 Moment Magnitudes and Stress Drops

In a second stage to the inversion theoretical t* values are computed by ray-tracing through the newly obtained Q=960 model. The site attenuation terms are also fixed as in Table 2.4 and a combined grid-search and Powell's search algorithm is performed for the source-corner frequency and the signal moment respectively by comparing the data with Equation 2. 1. After obtaining these values we can compute the seismic moment using Equation 2.4 and then the moment magnitude using Equation 2.6. For this computation we assume (correctly in this case) that the geometrical decay is given by 1/R; the free surface amplification F=2; the average radiation pattern $\Theta_{\lambda\phi}$ =0.55; v =3500m/s and ρ = 2800kgm^{-3}. The resulting moment magnitudes are shown in Figure 2.8. The inversion results are within 0.2 units of the correct magnitude, but tend to be overestimated rather than underestimated.

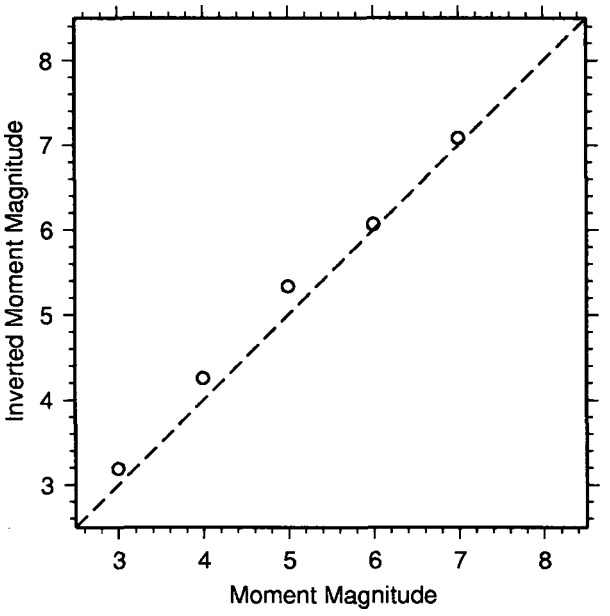

Figure 2.8: Plot of inverted moment magnitudes compared to the real magnitude used in the stochastic simulation from the inversion of M1.

Equations 2.5 and 2.7 are used to estimate the stress drop of these events assuming a circular rupture surface. Figure 2.9 illustrates that we are able to recover the average stress drop (1MPa) well, with a log-mean of 0.8MPa. The logarithm of the stress drop is given by:

$$\log(\Delta\sigma)=\log(0.76\text{MPa})\pm 0.06 \tag{2.13}$$

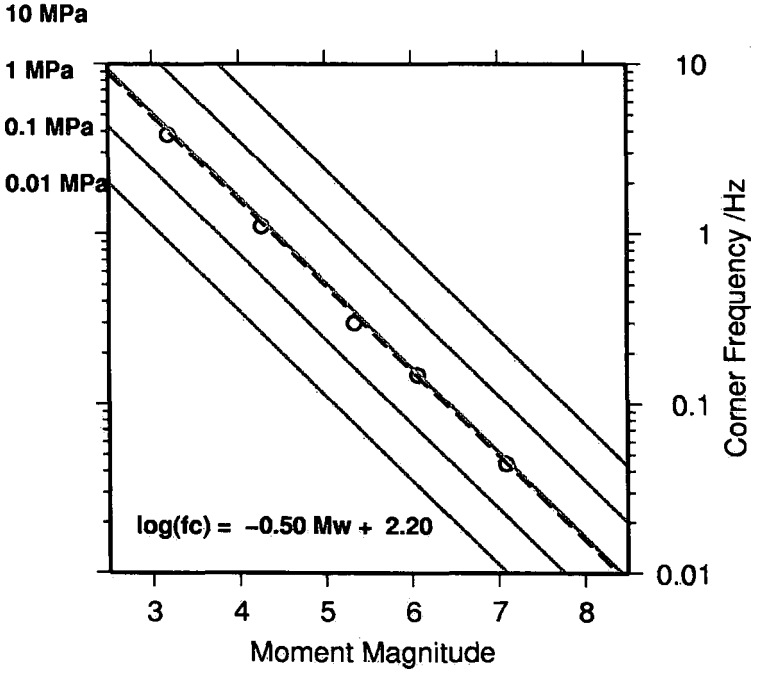

Figure 2.9: Plot of recovered source-corner frequency against recovered moment magnitude showing lines of constant stress-drop (grey) along with the mean stress drop (dashed line) from the inversion of M1.

Sensor	Hypocentral Distance (km)	Aj
1	10	1.17
2	20	1.11
3	30	1.06
4	40	1.02
5	50	1.01
6	60	1
7	70	0.98
8	80	0.97
9	90	0.97
10	100	0.96
11	120	0.95
12	150	0.91
13	200	0.92

Table 2.5: Recovered site amplification terms (A_j) from the inversion of M1. The input synthetic value was A_j=1.0.

Table 2.5 shows the inverted site amplification terms (A_j). We see that there is again a trend between the distance of the sensor from the source and the site amplification term (as with the site attenuation term (κ_j). These results show that there is some trade-off between the amplitude parameter and the attenuation parameter. However, for sensors further than 20km from the source we can find the correct A_j value to within around 10%. Figure 2.10 shows the frequency dependent site terms $a_j(f)$. The input was a flat response, i.e., $a_j(f)$=1. The recovered value is mostly correct. For the closer sensors there is a large scatter in the low frequencies (f<1Hz), and in general all the scatter in the term $a_j(f)$ increases as the frequency decreases. This is because there are comparatively fewer data points between 0.1 and 1Hz than between 1 and 10Hz (i.e., we are looking in the log-space but the data are equally spaced in the lin-space). As a result the inversion tends to favour fitting the higher frequencies at a cost of the lower frequencies.

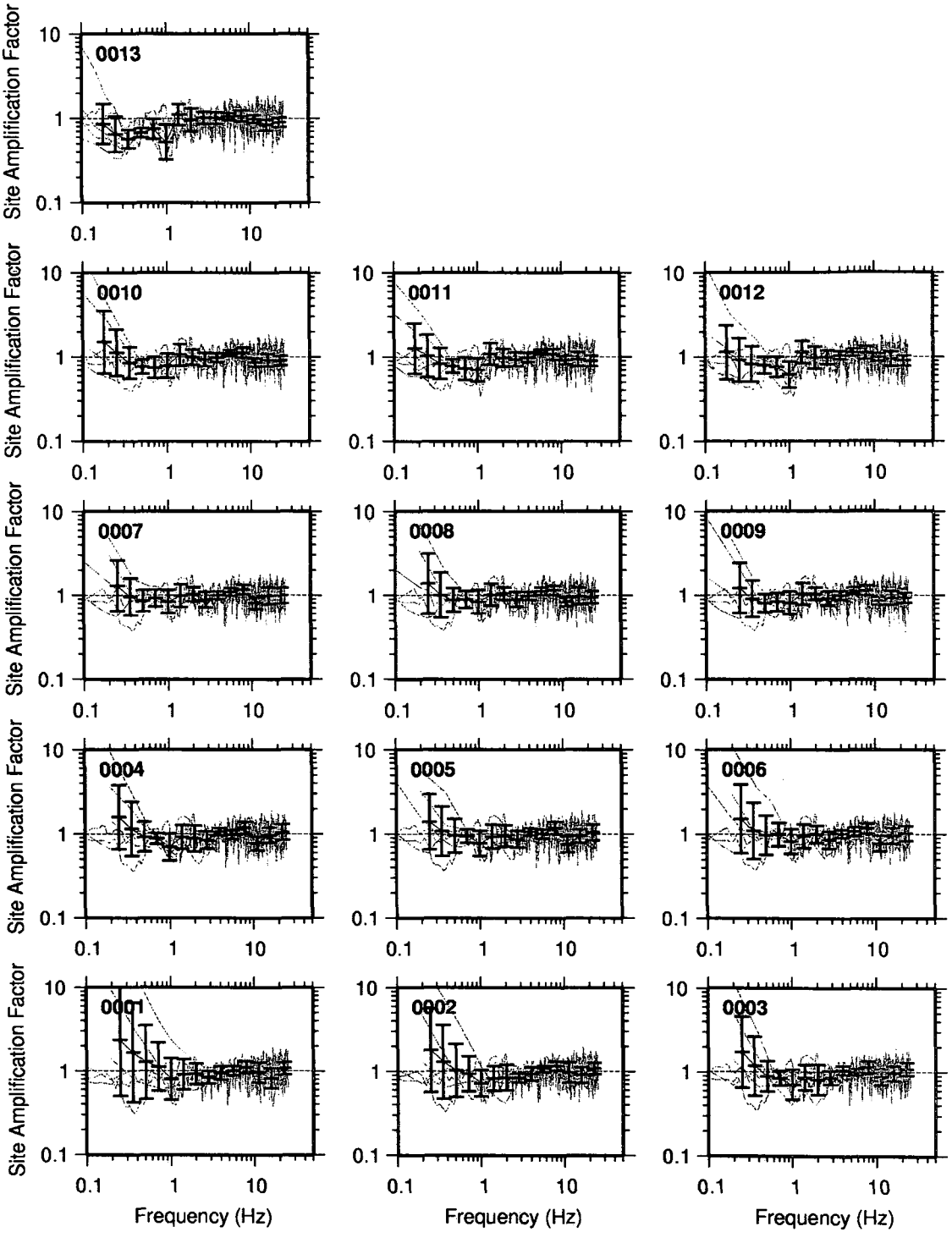

Figure 2.10: The site amplification term ($a_j(f)$) is shown for each of the 13 sensors used in the inversion of M1. The sensor number refers to the number used in tables 2.5 and 2.4.

The overall residuals from the inversion are shown in Figure 2.11.

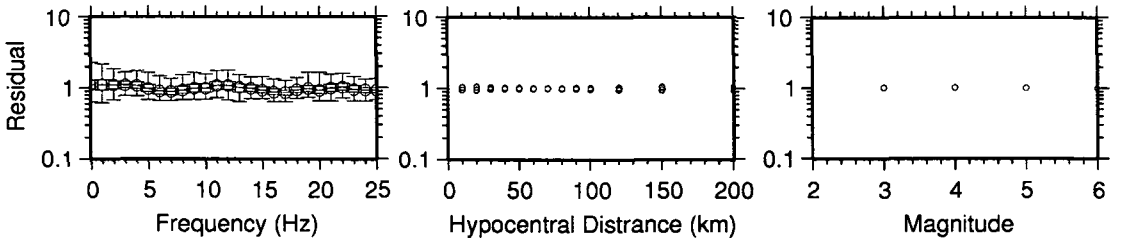

Figure 2.11: Residuals (data/model) of the complete inversion procedure of M1.For the plot against frequency the residuals are shown with the 63% and 95% confidence intervals (the box and whisker respectively).

2.3.5 Inclusion of high frequency site effects

In the method described in Chapter 1 we do not directly invert for the frequency dependent site response ($a_j(f)$). Instead we assume that it is given by the mean residual at each site (e.g., Figure 2.10). As the inversion attempts to minimise residuals for each event this may result in the site effects being mapped into other parameters. This results in a significant limitation to the method if, in reality, the site effects are not, on average, unity across the bandwidth used. We repeat the previous test, using synthetic data from the stochastic method of Boore (online), but this time we set the parameter aj(f)≠1 across the frequency bandwidth. We apply amplification to high frequencies by setting:

$$a_j(f) = \begin{matrix} 1.0 \\ 3.0 \end{matrix} \left| \begin{matrix} 0 \le f \le 10 \\ f \ge 15 \end{matrix} \right. \tag{2.14}$$

With linear interpolation between 10 and 15Hz (Figure 2.12). The other parameters are all as with M1. This model will be termed M2.

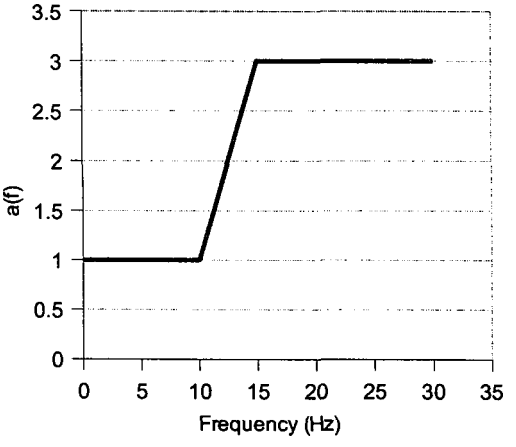

Figure 2.12: Site transfer term, $a_j(f)$, used in M2.

The resulting homogeneous Q model is 2500 (to 2 s.f.). This is far greater than the real value of 1000. Table 2.6 shows the inverted site attenuation terms (κ_j). Inverting M2 we found that the κ_j value decreases as the hypocentral distance increases up to 80km after which it increases in a pattern similar to the that seen when inverting M1. There is, however, a constant offset in κ_j between the results from M2 and M1. In this case all values are lower, with the magnitude of the extreme value (i.e., the minima) being larger. The scatter and resolution of κ_j is the same in both inversions. The negative values of κ_j reduce even further the attenuation applied to the spectra (Q=2500 is already low attenuation as attenuation is proportional to 1/Q).

Sensor	Hypocentral Distance (km)	κ (s)	±σ	Resolution
1	10	-0.001	0.002	1.00
2	20	-0.002	0.004	1.00
3	30	-0.004	0.006	0.99
4	40	-0.005	0.008	0.99
5	50	-0.006	0.010	0.98
6	60	-0.007	0.012	0.97
7	70	-0.008	0.014	0.96
8	80	-0.007	0.016	0.95
9	90	-0.006	0.018	0.93
10	100	-0.004	0.020	0.91
11	120	0.000	0.024	0.88
12	150	0.004	0.030	0.81
13	200	0.012	0.041	0.65

Table 2.6: Recovered station attenuation parameters (κ_j) from the inversion of M2. The distance from the source to the sensor is shown along with the standard error and the parameter resolution value. The input κ_j=0.000.

Figure 2.13 shows the inverted moment magnitudes and Figure 2.14 shows the source-corner frequency plotted against moment magnitude, indicating lines of constant stress drop.

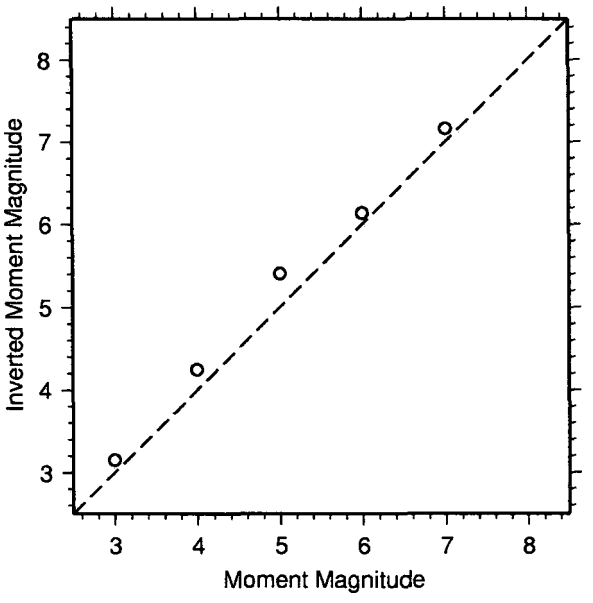

Figure 2.13: Plot of inverted moment magnitudes compared to the real magnitude used in the stochastic simulation of M2.

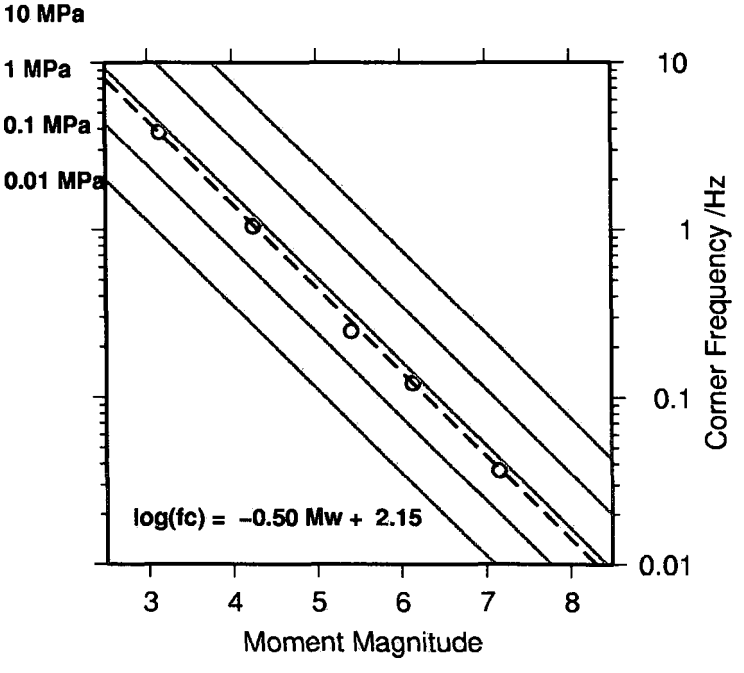

Figure 2.14: Plot of recovered source-corner frequency against recovered moment magnitude showing lines of constant stress-drop (grey) along with the mean stress drop (dashed line) using data simulated from M2.

There is no significant difference between the moment magnitudes recovered in this case (Figure 2.13) compared to the inversion of M1 (Figure 2.8). Figure 2.14 shows that the average recovered source corner-frequency is lower in this case than for M1 (Figure 2.9). As a result of this the average stress drop is lower for M2. The logarithm of the stress drop in

this case is given by:

$$\log(\Delta\sigma) = \log(0.64\text{MPa}) \pm 0.11 \quad . \tag{2.15}$$

Which is around 25% lower than the recovered value from M1 and 36% lower than the input of 1MPa.

Sensor	Hypocentral Distance (km)	Aj
1	10	1.17
2	20	1.11
3	30	1.06
4	40	1.02
5	50	1.01
6	60	1
7	70	0.98
8	80	0.97
9	90	0.97
10	100	0.96
11	120	0.95
12	150	0.91
13	200	0.92

Table 2.7: Recovered site amplification terms (A_j) from the inversion of M2. The input synthetic value was A_j=1.0.

Comparing Tables 2.7 and 2.5 we see that, as with the other frequency independent parameter, the moment magnitude, the site amplification term (A_j) is no different in the case of inverting M2 (with site effects) or M1 (without site effects). The inverted site terms ($a_j(f)$) are shown in Figure 2.15. The target function is given by Equation 2.14. It is clear that the correct form has not been recovered. We do see an increase in $a_j(f)$ starting at around 10Hz, but this increase is to approximately 2, and beyond this $a_j(f)$ reduces with frequency: i.e., there is a spike at the inflection point of the input $a_j(f)$ as shown in Figure 2.12. This suggests that the attenuation term t* is able to compensate for the high frequency site effect, and twists the residual function back towards a mean of 1.0. However, around the inflection point the model struggles to compensate for the site response, as such we observe only this part of the true site response in the inversion residuals. A discussion of this effect is given toward the end of this chapter.

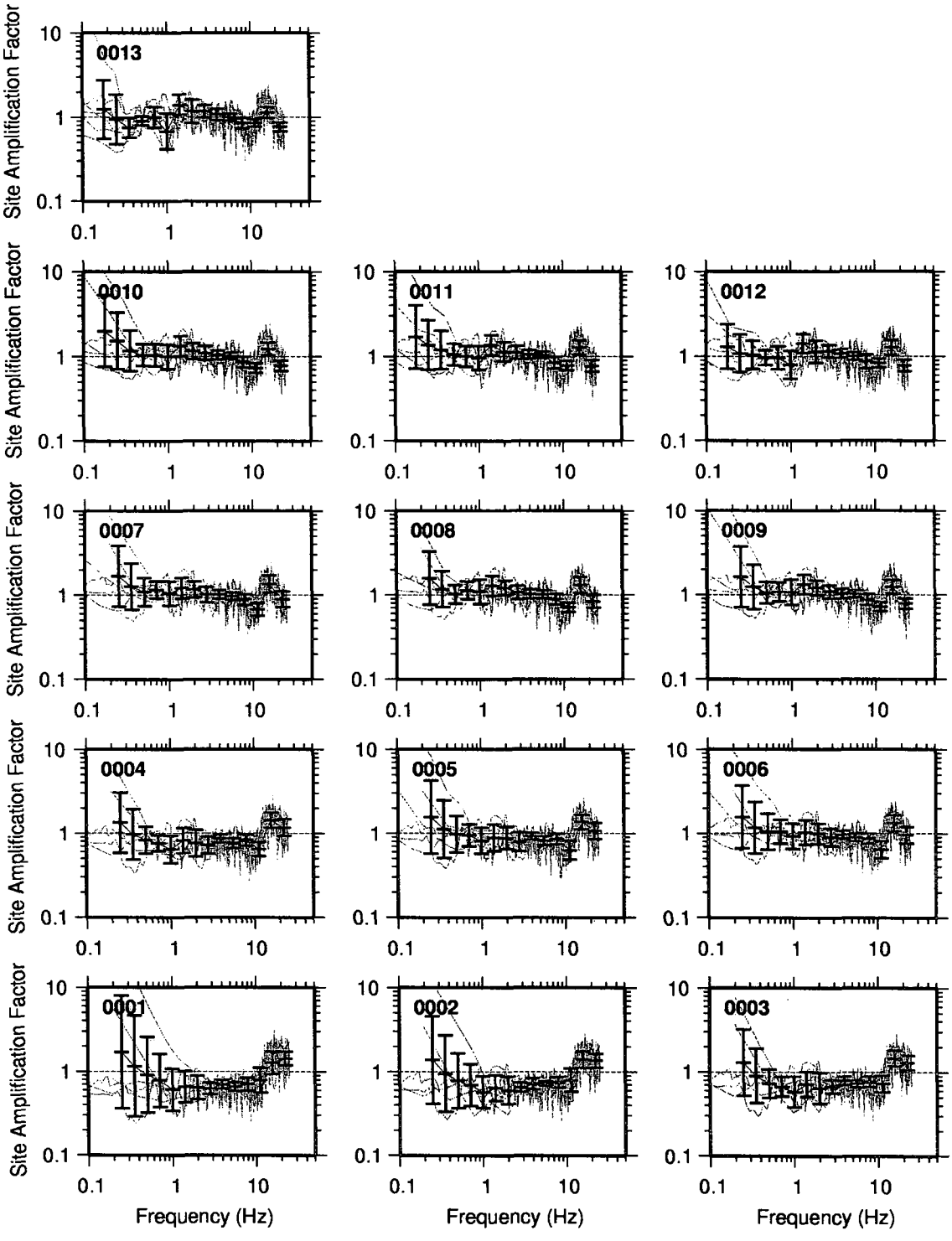

Figure 2.15: The site amplification term ($a_j(f)$) is shown for each of the 13 sensors used. The sensor numbers refer to the numbers used in e.g., Table 2.6. The input $a_j(f)$ is given by Equation 2.14.

The overall fit residuals, taking into account the average site response a(f), are given in Figure 2.16.

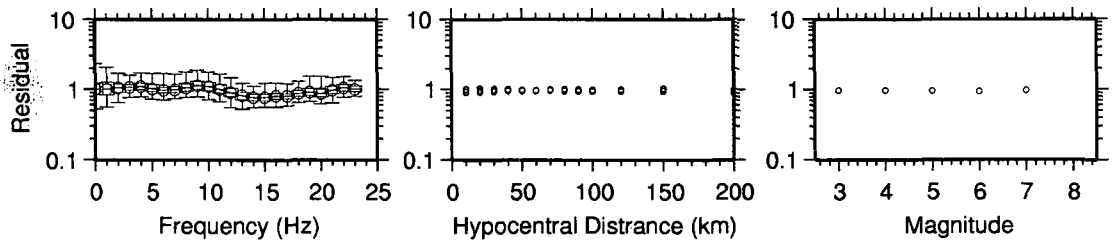

Figure 2.16: Overall residuals from the inversion of M2. For the plot against frequency the residuals are shown with the 63% and 95% confidence intervals (the box and whisker respectively).

2.3.6 Summary of M2

By comparing the results from inverting M1 and M2, where the only difference was the inclusion of a frequency dependent site term in M2, we can observe how the results change, and obtain an indication of how our real data may be influenced by strong site amplification.

The inclusion of Equation 2.14 in M2 led to negligible differences in the inverted signal moments, the spectral amplitude parameter. As such, the resulting decomposition into moment magnitudes and site amplification (A_j not $a_j(f)$) was independent of the site effect. However, the recovered Q parameter was severely influenced by the site term, with Q=2500 for M2: an increase of around 150%. In addition the source corner frequency, and the stress drop computed from this, were underestimated to a greater degree than the inversion of M1: a decrease of roughly 25% in the average stress drop was induced by the inclusion of the site term in M2. The function $a_j(f)$ is not correctly recovered at distances beyond 30km from the source, although we see spike at approximately the transition from 1 to 3 in input $a_j(f)$. There is a trade off between the parameter $a_j(f)$, Q, and to lesser degree, the source corner frequency. At distances closer than 30km the shape of $a_j(f)$ is recovered, as the attenuation due to Q at these distances is insufficient to trade-off with the site term. $a_j(f)$ from these close recordings is constantly underestimated, but the combined site term $A_j\, a_j(f)$ is correct as A_j is also overestimated at short distances.

2.3.7 Inclusion of low and high frequency site effects

In a second case we apply amplification over both low and high frequencies by setting:

$$a_j(f) = \begin{matrix} 1.0 \\ 3.0 \end{matrix} \begin{matrix} \Big\} 0 \le f \le 0.5 \\ \Big\} f \ge 1 \end{matrix} \qquad (2.16)$$

with linear interpolation between 0.5 and 1Hz (Figure 2.17). The other parameters are all as with M1. This model will be termed M3.

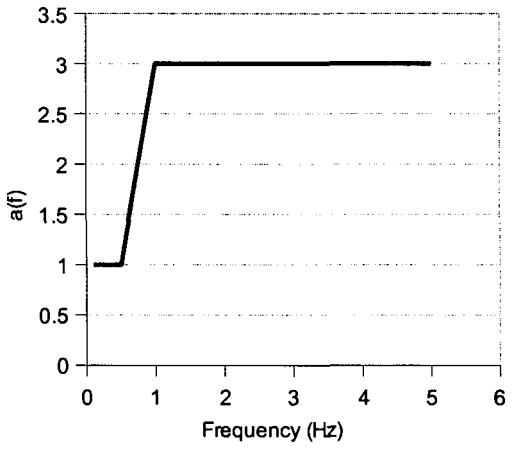

Figure 2.17: Site transfer term, $a_i(f)$, used in M3.

Inverting the synthetic data produced using this model we are able to correctly recover the Q=1000 value. This is not surprising, as we are now applying frequency dependent amplification to lower frequencies. The higher frequencies are all amplified by the same degree. As a result there is no change to the curvature of the spectrum at these higher frequencies and the recovered Q is the same as for M1. However, The signal moment and the source corner frequency are far more sensitive to this new $a_i(f)$ term.

The moment magnitudes recovered from M3 are plotted in Figure 2.18. The source corner frequency is plotted against moment magnitude for each event in Figure 2.19.

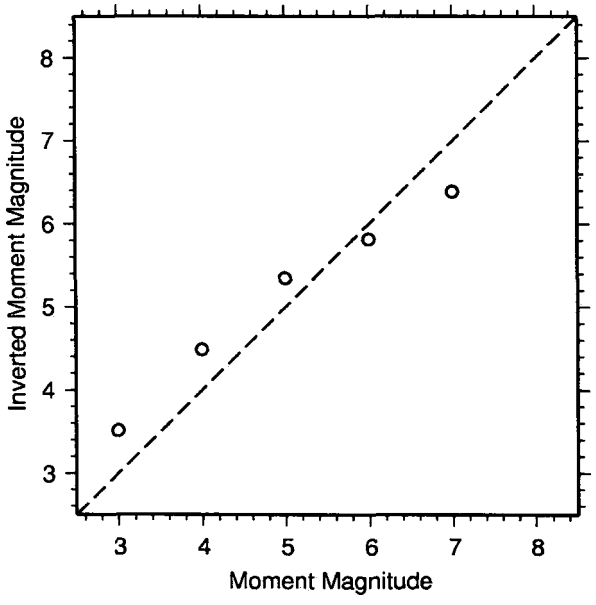

Figure 2.18: Plot of inverted moment magnitudes compared to the real magnitude used in the stochastic simulation of M3.

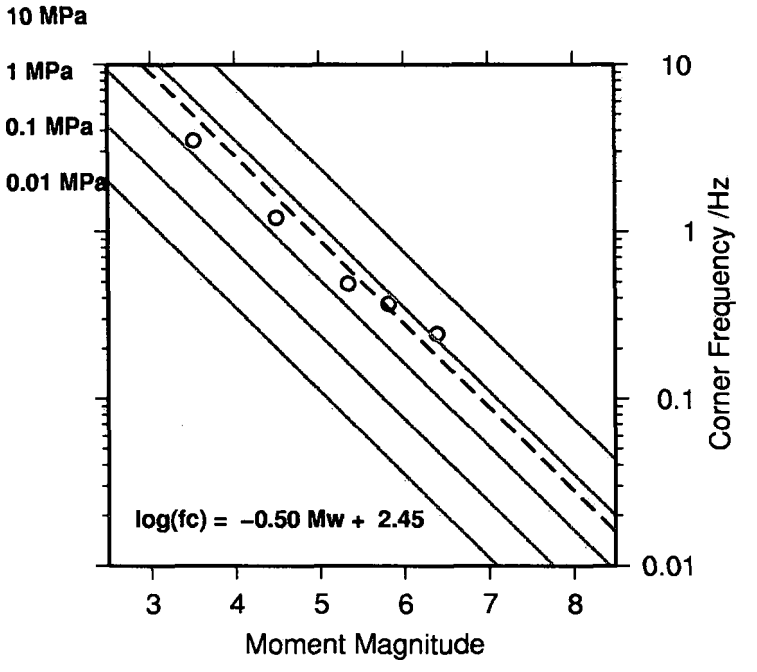

Figure 2.19: Plot of recovered source-corner frequency against recovered moment magnitude showing lines of constant stress-drop (grey) along with the mean stress drop (dashed line) using data simulated from M3.

The effect of the site term included in M3 is quite different from the effects seen in the inversion of M2 data. For the events with inverted corner-frequencies within the range affected by the site term (greater than 0.5Hz) the inverted moment magnitudes are overestimated and the source corner-frequencies are underestimated. For events with an inverted source corner-frequency of less than 0.5Hz, the moment magnitude is underestimated and the source corner frequency is overestimated. This leads to an increase in the average stress drop to 50bars, five times the input value.

Sensor	Hypocentral Distance (km)	Aj	κ (s)	K ±σ	κ Resolution
1	10	0.97	0.00	0.03	1.00
2	20	1.09	0.00	0.06	1.00
3	30	0.94	0.00	0.10	0.99
4	40	0.92	0.00	0.13	0.99
5	50	0.88	0.00	0.16	0.98
6	60	1.09	0.00	0.19	0.97
7	70	1.07	0.00	0.22	0.96
8	80	1.04	0.00	0.25	0.95
9	90	1.03	0.00	0.28	0.93
10	100	1.02	0.00	0.31	0.92
11	120	1.01	0.00	0.38	0.88
12	150	1.00	0.00	0.48	0.81
13	200	0.98	0.00	0.64	0.65

Table 2.8: Recovered station attenuation (κj) and amplification Aj parameters from the inversion of M3. The distance from the source to the sensor is shown along with the standard error and the parameter resolution value. The input κj=0.000 and Aj =1.0.

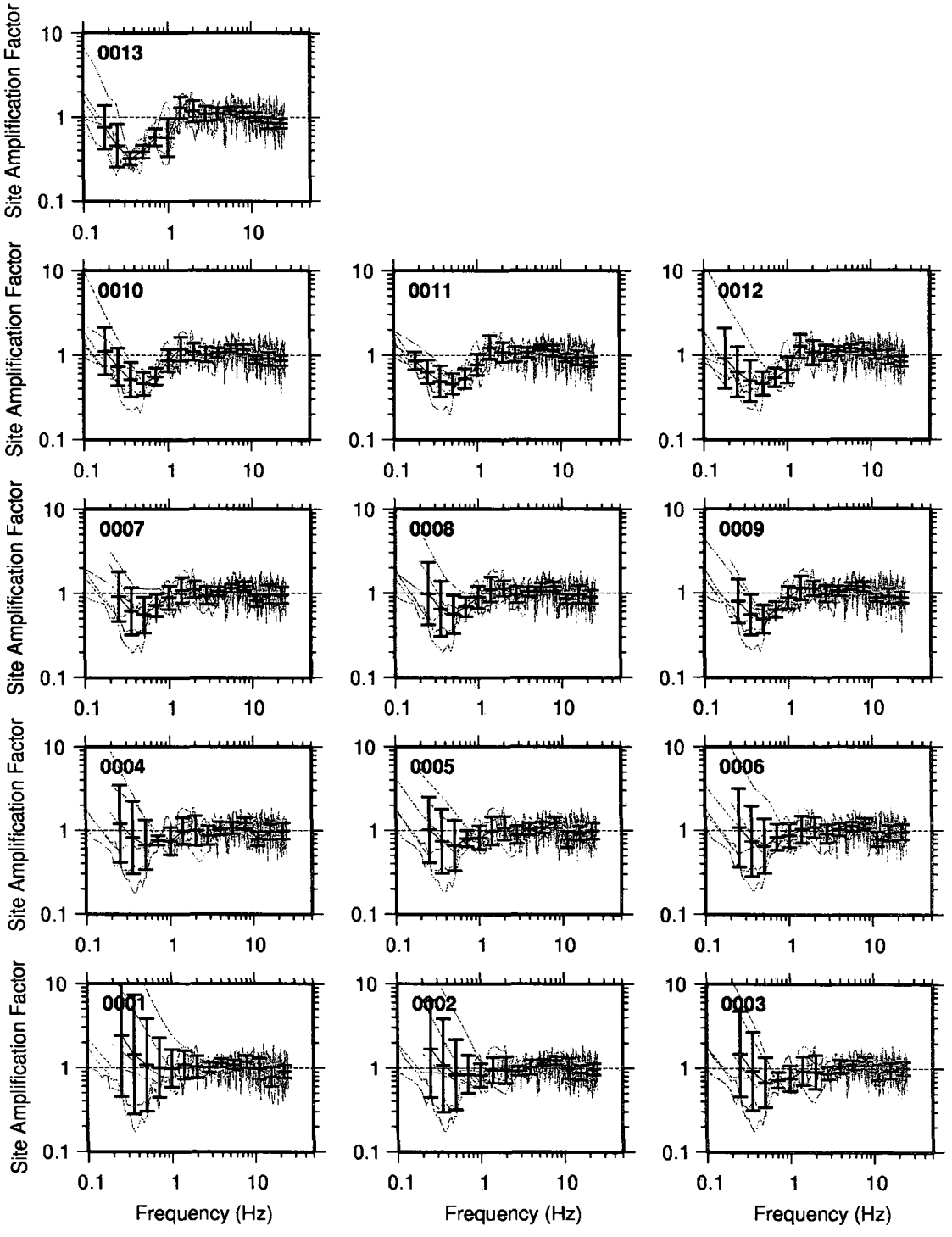

Figure 2.20: The site amplification term for M3 ($a_j(f)$) is shown for each of the 13 sensors used. The sensor numbers refer to the numbers used in e.g., Table 2.8. The input $a_j(f)$ is given by Equation 2.16.

The inverted site attenuation and amplification parameters (Table 2.8) are correctly recovered in this case. The site response terms, aj(f), are shown in Figure 2.20. From 0.5Hz the correct form of the function is recovered for records greater than 80km away from the source However, the amplitude of the response is incorrect by a factor of approximately 1/3. The closer observations give a relatively flat response across the bandwidth.

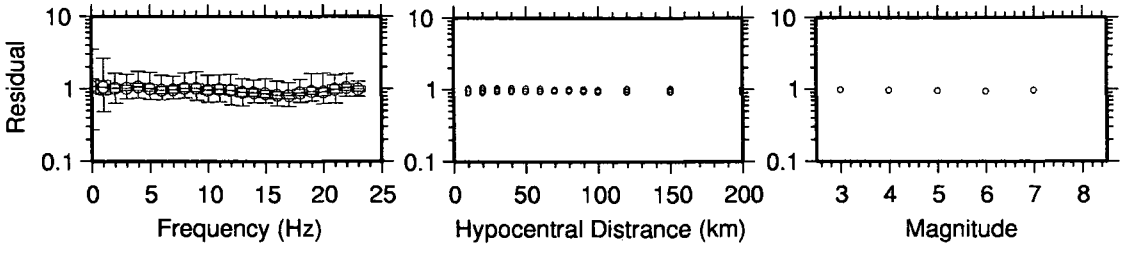

Figure 2.21: Overall residuals from the inversion of M3. For the plot against frequency the residuals are shown with the 63% and 95% confidence intervals (the box and whisker respectively).

The overall residuals are shown in Figure 2.21. The combination of parameters described previously are able to compensate for the inclusion of the site term to a good degree.

2.3.8 Combination of site effects

The final test in this chapter looks at the results of the inversion when a combination of site effects are applied such that the mean response $a_j(f)$ over all sensors is 1. The site response function used in M3 (Equation 2.16) is used in addition to the inverse of that function:

$$a_j(f) = \begin{matrix} 1.0 \\ 0.33 \end{matrix} \Bigg\} \begin{matrix} 0 \le f \le 0.5 \\ f \ge 1 \end{matrix} \tag{2.17}$$

with linear interpolation between 0.5 and 1Hz (Figure 2.22). Once again, the other parameters are all as for M1. This model will be termed M3. The two site response functions are applied alternately at sensors with increasing distance from the source. For instance, at sensor 1 $a_j(f)$ given by Equation 2.16 is applied; then for sensor 2 $a_j(f)$ given by Equation 2. 17 is applied; at sensor 3 $a_j(f)$ given by Equation 2.16 is applied, and so on.

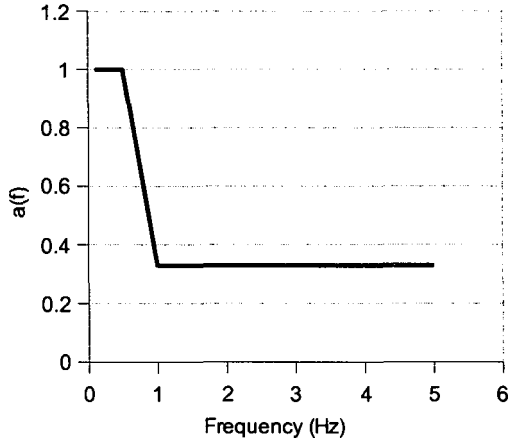

Figure 2.22: Site transfer term, $a_i(f)$, used in M3.

The recovered Q was 1040, close to the input of Q=1000. The M_w are close to the input values, but are all slightly overestimated by between 0.2 and 0.3 (Figure 2.23). There is no obvious trend to this discrepancy.

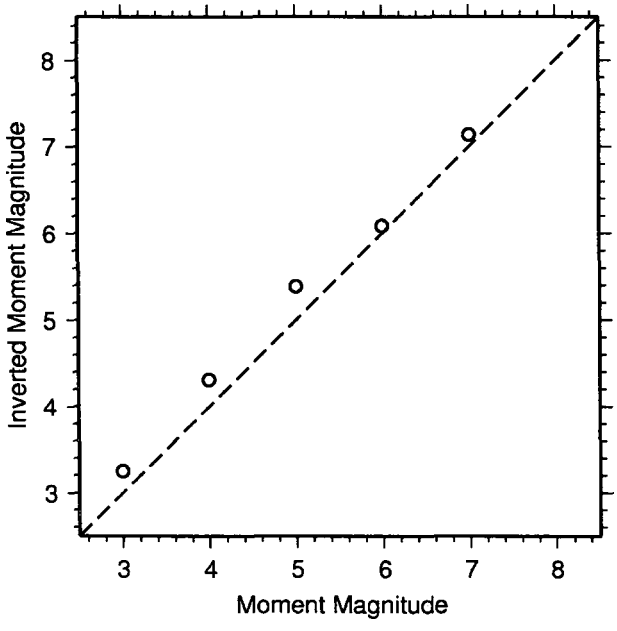

Figure 2.23: Plot of inverted moment magnitudes compared to the real magnitude used in the stochastic simulation of M4.

The stress drop recovered from the dataset M4 was:

$$\log(\Delta\sigma) = \log(0.67\text{MPa}) \pm 0.07 \quad . \tag{2.18}$$

This is the only 0.1MPa lower than was recovered for the dataset M1, where $a_i(f)=1.0$. A plot of corner frequency against M_w is shown in Figure 2.24. It is clear from this plot that the

source corner frequency is still somewhat underestimated, as the points should lie on the line of constant stress drop of 10MPa. This may be emphasised by the grid-search spacing of 5% for the source corner frequency, which equates to a spacing of approximately 15% for the stress drop.

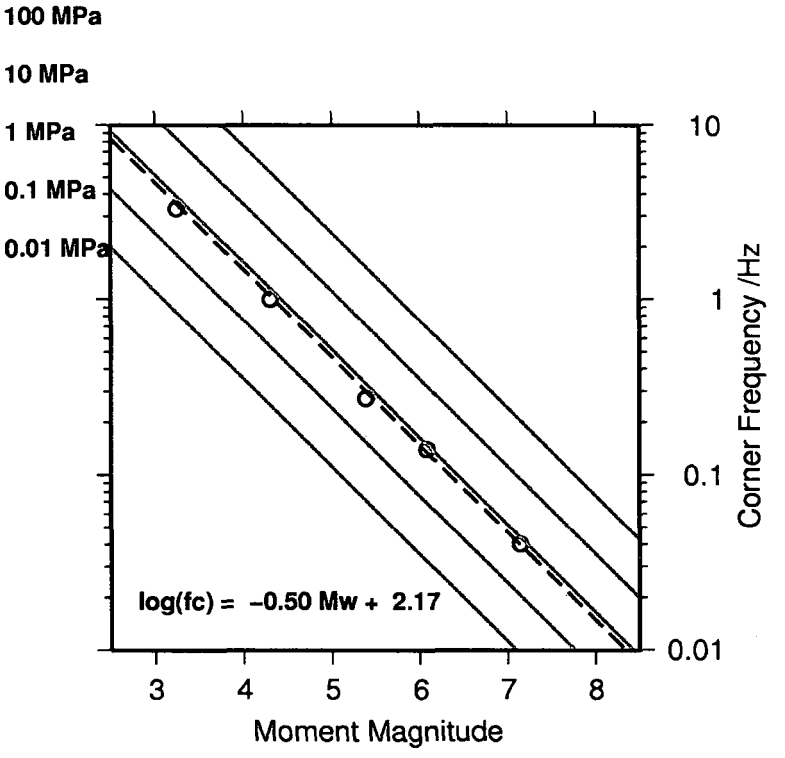

Figure 2.24: Plot of recovered source-corner frequency against recovered moment magnitude showing lines of constant stress-drop (grey) along with the mean stress drop (dashed line) using data simulated from M4.

Figure 2.25 shows that, in this case the site response values are very well resolved. The alternating site response can be made out with sensors 1, 3, 5, etc. amplifying the low frequency and sensors 2, 4, 6, etc. attenuating the low frequencies. $a_i(f)$ for the sensors that amplify the low frequencies are much better resolved below 0.5Hz than for the sensors that attenuate the low frequency.

This test has shown that the method performs as expected, and not far worse than for the model (M1) with a(f)=1.0, as long as product of the site response functions recording a particular event cancel out. Figure 2.26 shows that the residuals have no dependency on magnitude, distance or frequency.

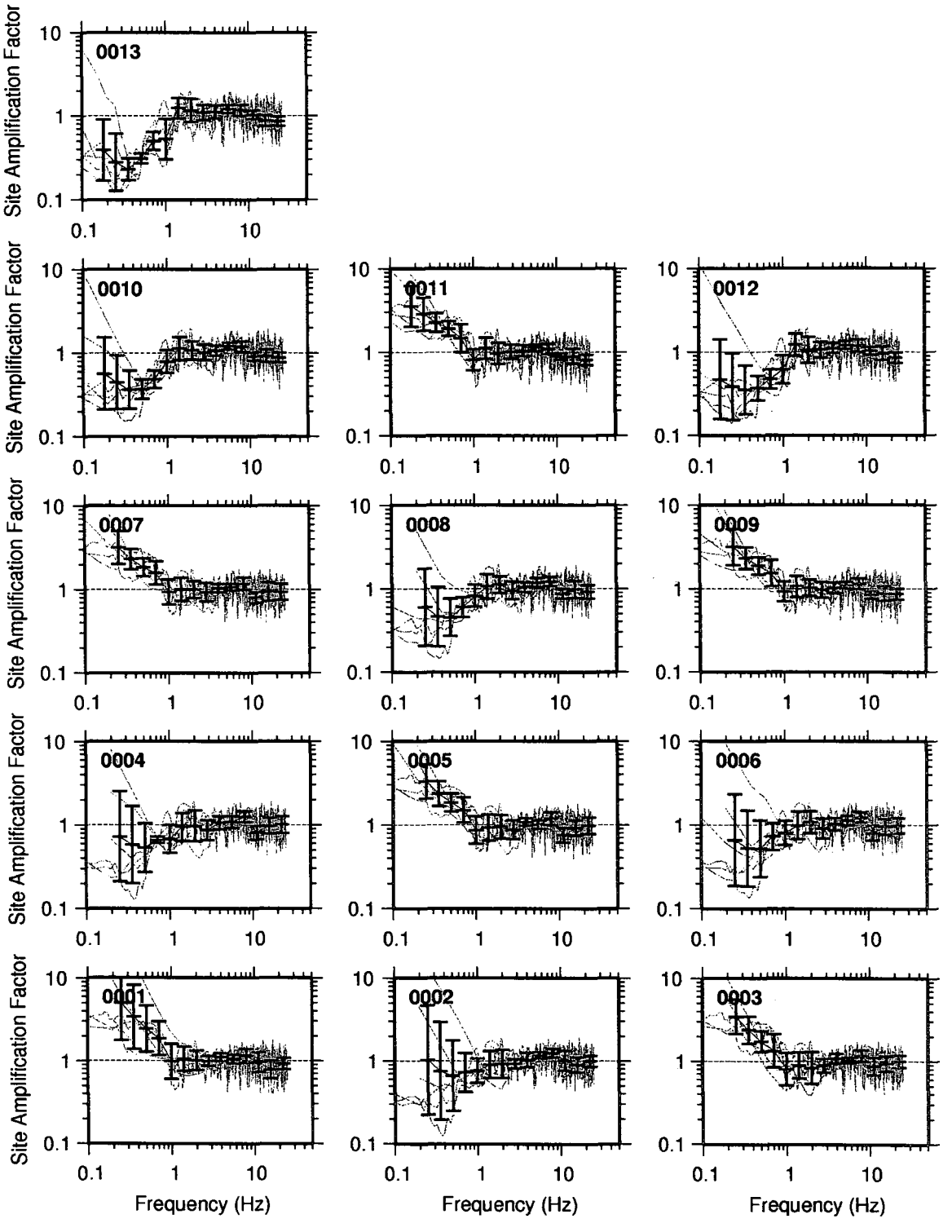

Figure 2.25: The site amplification term for M4 ($a_j(f)$) is shown for each of the 13 sensors used. The sensor numbers refer to the numbers used in the tables (e.g., Table 2.6). The input $a_j(f)$ is given by Equation 2.16 for odd sensor numbers and by Equation 2.17 for even sensor numbers.

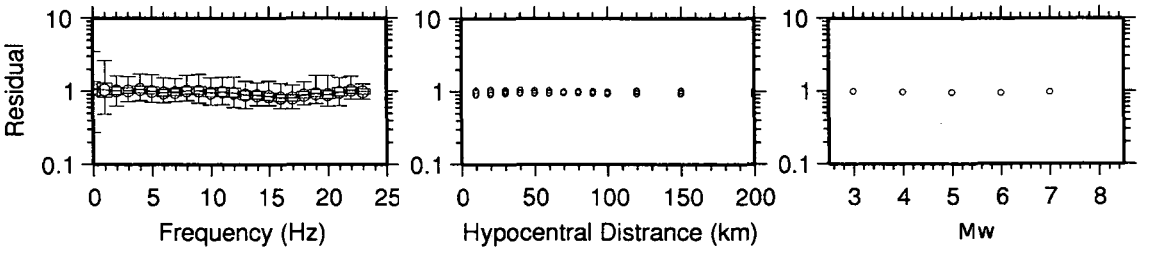

Figure 2.26: Overall residuals from the inversion of M4. For the plot against frequency the residuals are shown with the 63% and 95% confidence intervals (the box and whisker respectively).

2.4 Discussion and Conclusion

It has been shown that the method described in Chapter 1 was able to correctly recover the moment magnitude of an earthquake simulated using finite-difference modelling. The stress drops were similar to those obtained from equations based on the fault area but not the 'true' stress drop. This showed that the point source simplification of the source does not affect the method and that the Brune (1970, 1971) instantaneous slip model was adequate to model the finite source, but the inverted stress parameter should be treated as an underestimation of the true stress drop. We found that the maximum frequency that we could model and still obtain realistic results around 1.6Hz, which is consistent with the limitations of the source modelling method.

The second set of synthetic tests covered in this chapter used a stochastic method to produce synthetic seismograms. When not including any site response the recovered values were close to the input values. There was a small trade off between the site attenuation parameter κ_j and the site amplification value A_j. The moment magnitudes were correctly recovered while the Brune stress drop was underestimated by around 15%. It should be noted that the input stress drop in the stochastic model should be correctly recovered as it is not the same as the true stress drop used in the finite difference model. When including a site response in the modelled seismogram the inversion could not correctly recover the spectral parameters or the site response. This problem arises due to neglecting the site response in the actual inversion, and simply assigning the some of the residuals of the inversion to the site response, $a_j(f)$. If, as in this simple case, the site response is not on average flat (i.e., $a(f)=1.0$) over all recordings of a particular event, or the data distribution is not ideal (e.g., a completely random source-recorder distribution) then the site response is mapped into the other parameters in order to minimise the inversion residuals. As some of the site response is now already accounted for in the inversion parameters, the recovered site response is also affected: it tends to be flattened out. Unfortunately, an ideal source-receiver distribution may not be available. For instance, particular events are recorded well only at particular sites (i.e., between 20 and a few 100 km). As such, in this case it is likely that the trade-off between the site effect, $a_j(f)$, and the source and site parameters may influence results. The effects on the results are largely dependent on the true form of the site

response function. For instance, for a(f) amplifying high frequencies (e.g., greater than 10Hz) we find that Q is most affected, with the amplitude parameters largely unaffected. Whereas for a(f) amplifying low frequencies the Q value is unaffected, but the stress drop and the moment magnitude are both affected significantly. We tested the hypothesis that if the product of the site response over all recordings of an earthquake was flat, the inversion results would be unaffected, or comparable to the results from M1. We found that overall the moment magnitude was correctly recovered, albeit a little too high. The stress drop was lower than was recovered for M1, but Q was comparably close to the input value.

By looking at the site response computed using each of the synthetic datasets we can relate the change in the attenuation or source parameters to how well the site term is recovered. This is because the method used will always attempt to minimise the residuals over the entire frequency bandwidth: sometimes at the cost of losing the site response information which, in our model, are stored in the residuals. In cases where the product of site responses over all records of an earthquake was not unity (e.g., M2 or M3), the model was able to 'twist' the site response function in order to give flat residuals over the entire, or most of the, frequency bandwidth. It did this by changing a combination of the attenuation parameters leading to incorrect results. However, when the site response is still present, but has an average of unity over all records (e.g., M4) the model cannot pivot the site response functions to leave flat residuals. This is due to the way in which the inversion is set up: it is no longer trivial to change the attenuation parameters to correct for the site response, as it is no longer has a common form across all observations.

An example of a velocity spectrum from M1 (with $a_j(f)=1.0$), and M3 (with $a_j(f)$ defined by Equation 2.16) is shown in Figure 2.27. The spectra share exactly the same source and path parameters, the only difference being $a_j(f)$. The response function recovered by dividing M3 by M1 is also shown in Figure 2.27.

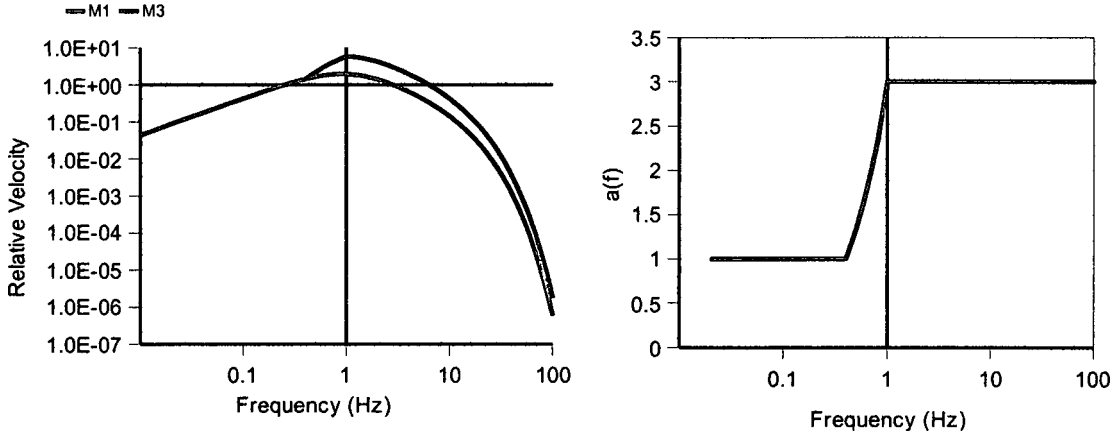

Figure 2.27: Left: example velocity spectra from M1 and M3; right: site response, $a_j(f)$, used in M3.

Ideally the inversion should find the same path attenuation and source parameters for both M1 and M3, and the residuals would reconstruct the site term $a_j(f)$. However, it has been shown that this is not always the case if the product of $a_j(f)$ across all records of one event is not 1. If, for instance we obtain only one record of this particular earthquake and it has the form of M3 the inversion method will find a solution similar to that shown in Figure 2.28. In this case the source corner frequency was increased by 70%, the other path attenuation parameters were correctly recovered. The resulting residuals, which define $a_j(f)$ show a spike (Figure 2.28) around the inflection of the true response function as shown in Figure 2.27. By increasing f_c the right hand limb of the response function is twisted back towards 1.

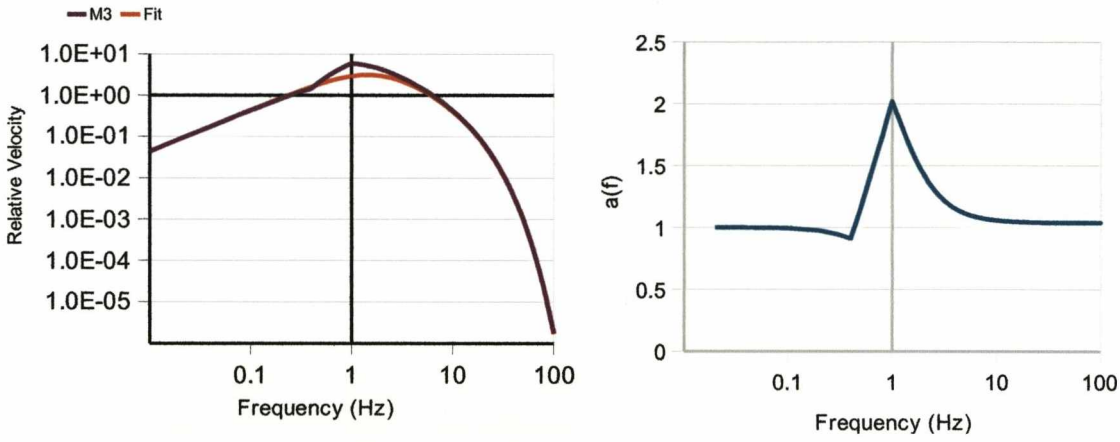

Figure 2.28: Left: example velocity spectra from M3 and an example fit using an artificially high f_c; right: resulting site response, $a_j(f)$, found for M3.

This is clearly an extreme case, as in reality we obtain numerous records of one earthquake at different locations. We must, however, use caution when interpreting the response function we obtain, as it our results may be influenced by a common form to the true site response function.

Chapter 3

3.1 Preface to Chapter 3

This chapter is an edited version of a paper accepted for publication in the Bulletin of the Seismological Society of America. Due to the structure of this thesis, repetition is sometimes unavoidable, as each chapter comprising of a scientific paper is a self-contained article. The motivation of this chapter was to develop attenuation relations for Great Britain. Attenuation relations are a critical component of seismic hazard analysis, which, in turn, is a key factor in mitigating seismic risk for both safety-critical structures such as biological-, chemical-, or nuclear facilities; and commercial or residential structures. The current energy review being carried out in the U.K. and the favoured expansion of nuclear power (Department of Trade and Industry, 2006, 2007) has renewed interest in seismic hazard in this region. Attenuation relations have been derived for Great Britain in the past (Principia Mechanica Ltd., 1982, 1988). However, these are now fairly outdated, and were based on world-wide earthquakes from similar (i.e., relatively stable) tectonic settings. The use of earthquake data from areas outside the study region is a common way of approaching the problem of predicting strong ground-motion. As previously discussed, strong-ground motion records are particularly sparse in regions of low to moderate seismicity due to, for instance, the calibration of older short period instruments and overall lack of strong-motion sensors. This problem is exasperated due to the long time-scales over which to mitigate losses (e.g., of the order several tens or even hundreds of years). Seismic recordings are, in this perspective, in their infancy. As a result we are unlikely to have a recordings of significantly strong ground motion in a study area such as the UK. Let alone numerous recordings required to perform a stable regression of the data. In recent years a significant amount of research has been directed towards this extrapolation problem. Propagating attenuation relations beyond our data constraints has been shown to be problematic when the error on this prediction is considered (e.g., Bommer 2006, Bommer et al., 2007). Instead of blind propagation, stronger ground motion data are 'imported' into the datasets, constraining the regression at higher levels of ground-motion.

The validity of attenuation relations derived from statistical regression is typically defined by the parameter range used in the regression. For instance, if the data used in the regression only included data from earthquakes of magnitude up to 6, the resulting relation is assumed to be valid for predicting ground-shaking from earthquakes of magnitude up to 6. Similarly for distance, depth and so on. This is due to the aleatory and epistemic uncertainty in the measure of ground-shaking leading to non-unique solutions. This is, of course, common in numerous geophysical problems. Extrapolation of a non-unique, statistical regression solution beyond the parameter range used in its derivation is risky. Therefore whilst a model will fit the data used in its derivation, the introduction of new data, particularly data from outside the valid parameter range (for instance higher magnitude earthquakes) will not necessarily fit the same model. Research has been undertaken to reduce the influence of

this scatter on the regression and resulting loss estimation, for instance, in its separation into aleatory variability and epistemic uncertainty (e.g., Toro et al., 1997) and inter- and intra-earthquake variability (e.g., Bommer and Crowley, 2006).

The aim of this project, however, is to overcome the limitations in the statistical regression model by changing the way in which the attenuation relations are formulated; from an empirically derived statistical model, to a model built from basic principles of seismology using data recorded explicitly in the study region, in this case Great Britain. In order to do this we use a stochastic simulation method (e.g., Boore, 2003) to simulate an earthquake of any size, for a given source mechanism. In order to be able to predict the ground motion at any given source and receiver location (the basis of an attenuation relation) we then need to know how the energy produced at the earthquakes source will propagate through the crust. The derivation and earthquake source model for Great Britain, and the properties of the underlying crust, are the subject of the following paper: The Acquisition of Source, Path and Site Effects from Micro-earthquake Recordings using Q Tomography: Application to the UK (Edwards et al. 2008). In the subsequent chapters of this thesis we deal with the conversion to the response spectrum through a stochastic method. The results in the following chapter are a vital input parameter in this method.

The Acquisition of Source, Path and Site Effects from Micro-earthquake Recordings using Q Tomography: Application to the UK.

Benjamin Edwards[1], Andreas Rietbrock[1], Julian Bommer[2], and Brian Baptie[3].

1. Department of Earth and Ocean Sciences, The University of Liverpool, UK.
2. Department of Civil and Environmental Engineering, Imperial College London, UK.
3. British Geological Survey, Edinburgh, UK.

Key words: attenuation, geometrical spreading, site response, seismic hazard, United Kingdom, source parameters, stress drop, seismic moment, moment magnitude, Q tomography.

3.2 Abstract

Source, site and propagation parameters are inverted from a UK database of weak-motion events ($2.0 > M_L > 4.7$). This results in the complete spectral parameterisation of over 3200 velocimetric records of 273 events from the year 1992 to 2006. The S wave is extracted from the vertical records and is processed using a multi-taper Fourier transform. We initially use a non-linear least squares log-space optimisation to obtain estimates of the attenuation parameter for each spectrum. The estimates of t* are then used to geometrically constrain a depth-dependent Q model using a technique adapted from velocity tomography. We then invert for the remaining frequency-dependent parameters and a collective amplitude parameter from the velocity spectra whilst fixing the newly computed attenuation parameters based on ray tracing through our Q model. The resultant amplitude parameters are then split into source moments, apparent geometrical spreading and site correction factors. We find a frequency-independent, depth-dependent Q structure. A linear relationship proportional to $0.7M_L$ between moment magnitude (M_w) and local magnitude (M_L) is found in the range of 2 to $4.7M_L$. The majority of stress drops are found to range from the order of 0.1 to 10MPa, with no evidence for increasing stress drop with magnitude. A multiple segment apparent geometrical spreading model is found to best describe the amplitude decay with distance, accounting for factors such as geometrical spreading and scattering, along with multiple phase interference in the analysis window. Site response functions are found to broadly correlate with regional geology, mean amplification occurring in the Cenozoic sedimentary rock sites to the south east of England relative to the harder Palaeozoic rock sites of Wales and Scotland. We use a bootstrap analysis technique to analyse the dependence of our results on the data in order to estimate the variance of the results and check the robustness of different inversions. Synthetic spectra are also computed in order to obtain minimum

variance and bias of model parameters associated with the method. In applying a geometrical Q constraint, through the use of Q tomography, we find that the robustness of the results is significantly increased. A thorough analysis of the trade-offs involved in the inversion is performed using synthetic datasets. We find strong trade-offs between the parameters, but we are able to show that this covariance is reduced when adopting the Q-tomography approach.

3.3 Introduction

The acquisition of region-specific ground-motion prediction equations for regions of low seismicity is of great significance for the seismic design of safety-critical facilities, yet presents numerous difficulties. Ground-motion estimation equations are a means of estimating the propagation of energy, in terms of ground-motion amplitudes, or their derivatives, from a defined source to a study location. They are a key tool used in seismic hazard analysis. In regions of moderate or high seismicity, the availability of large strong-motion datasets is commonly utilised by directly solving the ground-motion equations. This is achieved through the minimisation of observed peak ground, or spectral, acceleration or velocity residuals relative to a theoretical model. In regions of low seismicity this approach is not viable due to the lack of data. Nevertheless the necessity to consider earthquakes with long recurrence intervals for the design of safety-critical structures requires consideration of strong ground-motion.

The lack of strong-motion records in regions of low-to-moderate magnitude seismicity, such as the UK, means that there is no trivial method to formulate ground-motion prediction equations which are also valid for strong ground-motion. There are two principal methods available; the first is the calibration of strong ground-motion equations from seismically active regions of the world (see Bommer et al., 2005; Scherbaum et al., 2005; and Cotton et al., 2006); the second is the use of synthetically-produced accelerograms generated with a stochastic method (e.g., Boore, 2003; Atkinson and Boore, 1995). Both methods may be calibrated using attenuation relations derived from weak motion records from the study region, or attenuation relations valid for similar tectonic regions. The common factor required is the derivation or adoption of valid attenuation relations. Due to the nature of what is essentially an extrapolation from the empirically-derived relations to higher magnitudes, it is essential that we can accurately define these relations, as any unaccounted errors will propagate through to the final ground-motion equations.

Previous ground-motion estimation equations for the UK (Principia Mechanica, 1982, 1988) were based on models formulated from the regression of global strong-motion data, where coverage is dominated by seismically active regions such as western North America and the Mediterranean. Carefully calibrated region-specific attenuation relations may reduce the

error by facilitating the conversion from ground-motion equations from seismically active regions to a UK-specific ground motion equation. Comparisons in other areas of Europe have shown significant differences between recorded ground-motion and models derived from global or non-local data. In such cases the ground-motions from small earthquakes are greatly over-predicted when using relations derived directly from strong ground-motion. These differences may be attributed to regional source scaling (e.g., Marin et al., 2004). It is also argued that the use of low-magnitude recordings to predict strong ground-motion and vice versa is problematic (Bommer, 2006; Bommer et al., 2007). However, examples of this occur when using the direct regression of response spectra, where no obvious physical properties can be examined.

The deconvolution of velocimetric data into physical parameters describing the propagation of energy is fraught with trade-offs and ambiguities. Most documented and significant is the Q (quality factor; a measure of attenuation) and f_c (source corner frequency) trade-off (e.g., Scherbaum, 1990; Bay et al., 2003). The strong trade-offs and ambiguities involved with the parameters lead to large errors if the inversion is unconstrained. Without addressing the errors and covariances in obtaining the attenuation relation we risk the propagation of significant uncertainty into any resulting predictive equations, and hence the assessment of seismic hazard. The methods applicable to deconvolve velocimetric data may be broadly categorised into those which use the decay of peak spectral amplitudes (e.g., Bay et al., 2003; Herrmann and Malagnini, 1996; Malagnini et al., 2000 and Ou and Herrmann, 1990), and those which use 'spectral shape' information (e.g., Haberland and Rietbrock, 2001; Rietbrock, 2001; Scherbaum, 1990). Methods which parameterise the change of spectral shape with epicentral distance suffer from poorly determined Q(f) when Q(f) is close to being proportional to frequency. However, methods which use the change of peak amplitudes with distance suffer more significantly from the trade-off between the source corner frequency and Q. In a new approach, an extension of the method of Rietbrock et al. (2001, 2004), the inversion of the Q structure is strongly decoupled from all parameters by using a combination of absolute amplitude and spectral shape information together with a tomographic technique to provide a geometrical constraint to the Q structure (e.g., Rietbrock, 2001; Rietbrock et al., 2004; Scherbaum, 1990). The application of a 'velocity tomography' style technique potentially, dependent on data coverage, allows for a more realistic three-dimensional Q structure (e.g., Rietbrock, 2001; Salah and Zhao, 2003; Lees and Lindley, 1994; and Tsumura et al., 2000), as opposed to the homogeneous half-space model commonly used.

3.4 Methodology

3.4.1 Method: Modelling the Fourier Spectrum

The Fourier velocity spectrum, $\Omega_{ij}(f,r)$, observed at a station, *j*, originating from an earthquake, *i*, can be represented as:

$$\Omega_{ij}(f,r)=2\pi f E_i(f)\cdot B_{ij}(f)\cdot S_{ij}(r)\cdot T_j(f)\cdot I_j(f)\,, \tag{3.1}$$

where f is the frequency, r is the hypocentral distance, $E_i(f)$ is the source model (the amplitude spectrum at the source), $B_{ij}(f)$ is the attenuation along the ray path, $S_{ij}(r)$ is the amplitude decay with distance or 'apparent geometrical spreading', $T_j(f)$ is the site transfer function at the station and $I_j(f)$ is the instrument response function. The source spectrum can be modelled by:

$$E(f)=\Psi_i\cdot\frac{1}{\left(1+\left(\frac{f}{f_c}\right)^{n\gamma}\right)^{1/n}}\,, \tag{3.2}$$

(Boatwright, 1978), where ψ_i is the long-period plateau value, f_c the source corner frequency, and n and γ are dependent on the source model and define the high-frequency decay at the source. For $\gamma=2.0$ and n=1.0, $E_i(f)$ is equivalent to the Brune ω^2 source model (Brune 1970, 1971). Assuming whole path attenuation, the attenuation along the path of propagation is:

$$B_{ij}(f)=e^{\left(\frac{(-\pi\cdot f\cdot t_{ij}^{*})}{f^{\alpha}}\right)}=e^{\left(-\pi\cdot f\cdot\frac{T_{ij}}{Q_{0ij}\cdot f^{\alpha}}\right)}\,, \tag{3.3}$$

where T_{ij} is the travel time, Q_{0ij} is the dimensionless quality factor (e.g., Rietbrock, 2001), f^{α} determines the frequency dependence of Q, and t_{ij}^{*} is the whole path attenuation operator defined by:

$$t_{ij}^{*}=\frac{T_{ij}}{Q_{0ij}}\,. \tag{3.4}$$

The apparent geometrical spreading function, which may include factors such as phase interference and dispersion, focusing or defocusing and scattering (see Atkinson and Mereu, 1992), is described by a piecewise function comprising of segments of constant exponential decay:

$$S_{ij}(r)=\begin{cases}\left(\frac{r_0}{r}\right)^{\lambda_1} & r\le r_1\\ S(r_1)\cdot\left(\frac{r_1}{r}\right)^{\lambda_2} & r_1\le r\le r_2\\ \vdots & \vdots\\ S(r_n)\cdot\left(\frac{r_n}{r}\right)^{\lambda_n} & r\ge r_n\end{cases}. \tag{3.5}$$

Where r_0 is a reference distance before which apparent geometrical spreading is assumed to be zero in order to take into account the complexity of the rupture surface, i.e., its deviation from a point source. $r_{i>0}$ are distances at which the geometrical spreading function segments. We may estimate r_0 using the source corner-frequency and assuming a circular rupture plane:

$$r_0=\frac{v_s k}{f_c}, \tag{3.6}$$

(Brune, 1970; 1971). For shear waves, v_s is the shear wave velocity at the source, and k is 0.37 (Brune, 1970; 1971). The parameter r_0 must also be used in the computation of the moment to restore the point source assumption; alternatively we may normalise the source size to r_0 =1km without affecting the moment or geometrical spreading value. The local site amplification is given by:

$$T_j(f)=A_j\cdot a_j(f)\cdot e^{\left(\frac{(-\pi\cdot f\cdot\kappa_j)}{f^{\alpha}}\right)}, \tag{3.7}$$

where A_j is a frequency-independent correction factor, and κ_j is a constant, site-related attenuation operator (e.g., Anderson and Hough, 1984) and $a_j(f)$ is the frequency dependent site amplification function which takes into account resonant frequencies due to the layered, fractured subsurface (e.g., Steidl et al., 1996). The data are carefully corrected for instrument response such that $I_j(f)=1.0$. Equation 3.1 can be rewritten using relations 2 to 7 such that the spectral amplitude at any given epicentral distance and frequency is given by:

$$\Omega_{ij}(f,r)=\Psi_i\cdot A_j\cdot S_{ij}(r)\cdot a_j(f)\cdot\frac{1}{\left(1+\left(\frac{f}{f_c}\right)^{n\gamma}\right)^{1/n}}\cdot e^{\left(-\pi\cdot f^{1-\alpha}\cdot\left(t_{ij}^{*}+\kappa_j\right)\right)} \quad (3.8)$$

The first three terms of which are frequency independent and we collectively term the signal moment, $\hat{\Omega}_{ij}(r)$:

$$\hat{\Omega}_{ij}(r)=\Psi_i\cdot A_j\cdot S_{ij}(r)\,. \quad (3.9)$$

We first solve Equation 3.8, obtaining the frequency dependent-parameters, and the combined amplitude parameter, the signal moment (Equation 3.9). The signal moments are subsequently inverted to solve for the individual distance-dependent amplification parameters. The complete inversion procedure is summarised by Figure 3.1, and described in detail below.

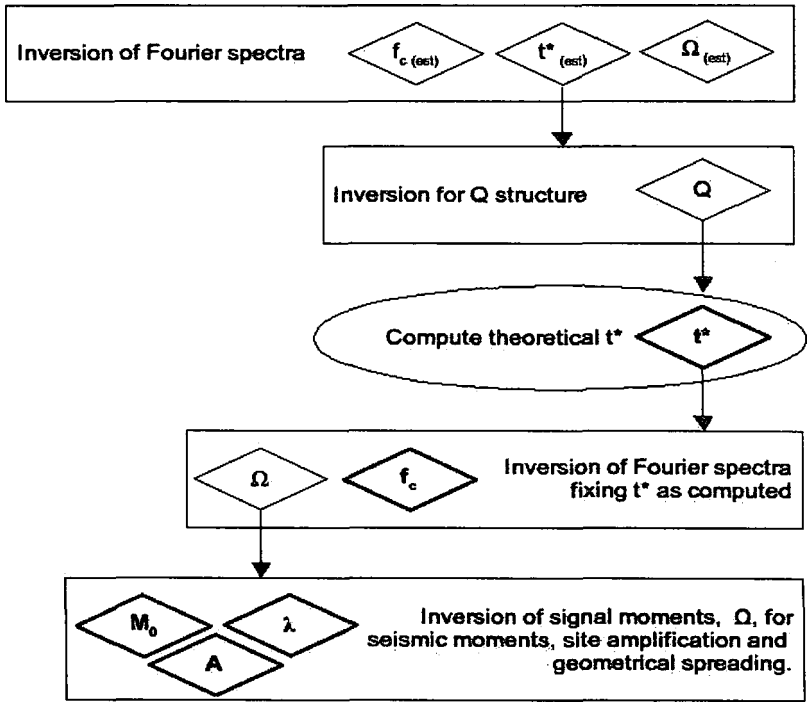

Figure 3.1: A flow chart of the method used. From top to bottom: (1) the initial spectral inversion; (2) t estimates are then used to construct a Q model using a tomographic method; (3) theoretical t* values are computed for each spectrum using the new Q model; (4) the spectral inversion is repeated, this time fixing the theoretical t* value; (5) finally the signal moment is decomposed into seismic moment, a site amplification term, and a geometrical decay value. Parameters in bold diamonds indicate the final values of each parameter.*

3.4.2 Method: Inversion of Attenuation, Source Corner Frequencies and Site Amplification

Following the approach of Rietbrock (1996, 2001) and Haberland and Rietbrock (2001), the shape of the Fourier spectrum from an observed event can be decomposed into a source corner frequency, individual signal moments and path dependent 'combined attenuation' operators, $t^*_{ij(comb)} = t^*_{ij} + \kappa_j$. For each event, a grid-search through the source corner frequency is combined with the optimisation of a residual fit function to extract the path-dependent parameters: the signal moments and attenuation operators. The inversion is initially only constrained by the assumption of a common source corner frequency, allowing t* to vary for each spectrum. The minimisation functions tested for use in the inversion were the following: absolute amplitude (L1) fit , given by:

$$\epsilon = \sum_{f=f_{start}}^{f_{end}} \frac{|\Omega^r_{ij}(f) - \Omega^m_{ij}(f)|}{\sigma_{ij}(f)} ; \tag{3.10}$$

the least-squares (L2) fit:

$$\epsilon = \sum_{f=f_{start}}^{f_{end}} \frac{\left(\Omega^r_{ij}(f) - \Omega^m_{ij}(f)\right)^2}{\sigma_{ij}(f)} ; \tag{3.11}$$

the maximum-likelihood fit (Rietbrock ,1996):

$$\epsilon = \sum_{f=f_{start}}^{f_{end}} \frac{\left(\Omega^r_{ij}(f) - \Omega^m_{ij}(f)\right)^2}{\sigma_{ij}(f) \cdot \Omega^m_{ij}(f)} ; \tag{3.12}$$

the log-space L1 fit:

$$\epsilon = \sum_{f=f_{start}}^{f_{end}} \frac{|\log(\Omega^r_{ij}(f)) - \log(\Omega^m_{ij}(f))|}{\log(\sigma_{ij}(f))} ; \tag{3.13}$$

and the log-space L2 fit:

$$\epsilon = \sum_{f=f_{start}}^{f_{end}} \frac{\left(\log(\Omega^r_{ij}(f)) - \log(\Omega^m_{ij}(f))\right)^2}{\log(\sigma_{ij}(f))} ; \tag{3.14}$$

where superscript *r* and *m* denote spectral values for the real data and the model respectively. Due to rigorous data processing and picking procedures, the variance $\sigma_{ii}(f)$

of each spectral data-point is assumed equal, and therefore irrelevant to the minimisation. The maximum-likelihood estimator (Rietbrock, 1996) effectively weights data from the high-frequency roll-off portion of the spectrum (i.e., beyond the source corner frequency), higher with respect to the same data in an L1 or L2 inversion. The L1 and L2 estimators give preference to fitting the higher amplitudes, from the plateau value to the source corner-frequency of the spectrum. This may lead to poor determination of Q. However, this 'automatic' weighting may correspond to the data reliability, as higher frequencies are more likely to be influenced by noise. The log-space minimisation functions have no such biases in fitting different sections of the spectrum; in this case we must be certain not to 'model the noise'. The choice of minimisation function is ultimately dependent on the data quality (in terms of both bandwidth and noise), along with the data type: displacement, velocity or acceleration spectra; as such synthetic or bootstrap tests are required to make an informed choice. In order to provide some stability to this initial inversion we require at least three observations of each event. While the Q tomography in the next stages will more fully stabilise the inversion this constraint helps to reduce the influence of single observations of earthquakes which may not be reliable.

3.4.3 Method: Q Tomography – Inversion for t*

In a second stage the obtained $t^*_{ij(comb)}$ estimates are used to reconstruct the Q structure in an approach similar to that used in seismic tomography. Using a known velocity model, the $t^*_{ij(comb)}$ value depends only on the Q^{-1} along the ray-path:

$$t^*_{ij(comb)} = \int_{raypath} \frac{1}{Q(x,y,z)\cdot v(x,y,z)} ds + \kappa_j \qquad (3.15)$$

(e.g., Cormier, 1982). For the tomographic reconstruction we use an adaptation (Rietbrock, 1996, 2001) of the tomography programme SIMUL2000 (Eberhart-Phillips, 1993; Thurber, 1983, 1993). This is an iterative damped L2 inversion solving for Q along the raypaths and κ_j, the site-dependent attenuation operator. Damping is empirically selected as the value which gives best data variance with respect to keeping model variance low: i.e., somewhere on the 'knee' of the parabola for data variance plotted against model variance (Eberhart-Phillips, 1993). Damping is first selected for the Q model, without inverting for κ_j, then the Q model is fixed at the results obtained with the optimum damping value and we repeat the process to find the optimal κ_j damping and associated parameters. Using this approach we avoid the possibility of influencing the Q model by using the site-dependent κ_j, and instead choose for κ_j to fit the t* residuals. For an event to be included in the inversion we require at least four t* records of epicentral distance no greater than 500km. No preferential weighting is applied to records of differing hypocentral distance.

In a final stage, synthetic combined $t^*_{ij(syn)}$ operators are computed for every observed station-event path using the Q structure and κ_j operators obtained in the previous step. The synthetic $t^*_{ij(syn)}$ operators are then fixed in a subsequent spectral inversion which solves exclusively for common source corner frequencies and independent signal moments. The same data selection criteria (frequency range, distance range, etc.) as used in the first spectral inversion are applied.

The factorial residuals of the above inversions, given by:

$$\theta_{ij}(f)=\frac{\Omega^o_{ij}(f,r)}{\Omega^m_{ij}(f,r)}, \tag{3.16}$$

where the superscripts 'o' and 'm' denote observed and modelled amplitudes respectively, can be used to reconstruct the frequency-dependent site function by taking the geometrical mean of the factorial residuals at discrete frequencies over all events (i=1,2,3...n) at a specific station, 'j' (e.g., Scherbaum, 1990; Drouet et al., 2005):

$$a_j(f)=\left(\prod_{i=1}^{n}\theta_{ij}(f)\right)^{\frac{1}{n}}. \tag{3.17}$$

We must be careful in interpreting these site response functions, however, because they are essentially the residuals of the previous inversions and hence are more susceptible to noise and instrument calibration errors.

The method described above largely decouples the attenuation operators and source corner frequencies since the attenuation structure is constrained by many different source-receiver geometries (e.g., Scherbaum, 1990; Rietbrock, 2001), whilst the source corner frequencies are constrained with the assumption that a common source spectrum exists for multiple observations of the same event. The signal moments are used in a further inversion to separate the frequency-independent parameters: Ω_i, A_i and $S_{ij}(r)$.

3.4.4 Method: Inversion of Apparent Geometrical Spreading, Mean Site Corrections and Seismic Moments

The signal moments can be expressed as a linear function by taking the logarithm of Equation 3.9:

$$log[\hat{\Omega}_{ij}(r)]=log[\Omega_i]+log[A_j]+log[S_{ij}(r)]\ . \tag{3.18}$$

In matrix form this can be represented as:

$$\begin{bmatrix} m_{11} \\ m_{12} \\ \vdots \\ m_{1J} \\ m_{21} \\ \vdots \\ m_{2J} \\ \vdots \\ m_{3J} \\ \vdots \\ m_{IJ} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 1 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 0 & 0 & \cdots & 1 \\ 0 & 1 & 0 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & 0 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \vdots & 0 & 0 & \cdots & \vdots \\ 0 & 0 & 1 & \cdots & 0 & R'_1 & R'_2 & \cdots & R'_n & 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & R'_1 & R'_2 & \cdots & R'_n & 0 & 0 & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_I \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ a_1 \\ a_2 \\ \vdots \\ a_J \end{bmatrix}, \tag{3.19}$$

where $m_{ij}=log[\hat{\Omega}_{ij}(r)]$, $\omega_i=log[\Omega_i]$, $a_j=log[A_j]$, λ_n is the constant exponent of the nth section of the apparent geometrical spreading function, and

$$R'_k=\begin{cases} log\dfrac{r_{(k-1)}}{r} & r_{k-1}<r<r_k \\ log\dfrac{r_{(k-1)}}{r_k} & r\geq r_k \end{cases}, \ k = 1,2,3...n, \tag{3.20}$$

where $r_{k>0}$ refers to boundary distances of the kth section of the constant exponent apparent geometrical spreading function and r_0 is the fault plane radius. For simplicity Equation 3.19 may be expressed as:

$$\boldsymbol{s}=\boldsymbol{Gx} \tag{3.21}$$

As we are in the log-log space the least squares minimisation of Equation 3.21 is given by:

$$min\left(\sum_{n=1}^{N}\left(\log(\hat{\Omega}_n(r))-s_n\right)^2=\sum_{n=1}^{N}\left(\log(\hat{\Omega}_n(r))-E[\log(\hat{\Omega}_n(r))]\right)^2=\sum_{n=1}^{N}\left(\log\left(\frac{\hat{\Omega}_n(r)}{E(\hat{\Omega}_n(r))}\right)\right)^2\right) \tag{3.22}$$

hence the solution to Equation 3.21 is a fractional minimisation of the actual and computed signal moments (Equation 3.22). This prevents domination of the inversion by large magnitude events. We solve Equation 3.22 using a 'least squares' (L2) minimisation with a

singular value decomposition (SVD) algorithm (Press et al., 1997; Pozo, 2004). The inversion is constrained by assuming:

$$\sum_{j=1}^{J} \log(A_j) = 0 \ , \tag{3.23}$$

such that the parameter A_j is defined as an amplification relative to the network average. This definition is taken due to the lack of detailed site responses for use as reference sites (e.g., Steidl et al., 1996).

The apparent geometrical spreading functions are initially constructed based upon a visual inspection of the amplitude distribution with respect to distance from the source and the time from the S-wave arrival (Figure 3.2), along with theoretical calculations for the arrival of new phases into the defined S-window.

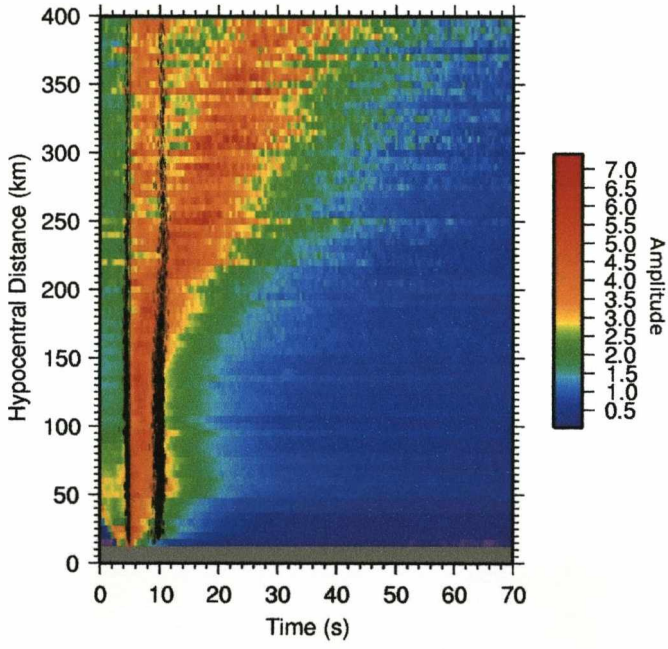

Figure 3.2: A plot of the average absolute amplitude at different hypocentral distances and time. The time axis is normalised such that the S wave onset is always at 5 seconds. The amplitude scale is normalised relative to each hypocentral distance (the effect of geometrical spreading is removed). The black lines indicate start and end of windows used in the spectral inversion.

These suggest that the first observations of S_n and L_g are around twice the crustal thickness and 130km respectively (e.g., Atkinson and Mereu, 1992). A grid-search is then performed through the possible combinations of boundary distances. Data constraints are applied to prevent poorly determined parameters: the number of observations of one event, and the overall number of observations at a station must both be greater than or equal to 5. Finally,

seismic moments can be calculated using the following equation:

$$M_i = \frac{\Omega_i \cdot v^3 \cdot \rho \cdot r_0 \cdot 4\pi}{\xi F \Theta_{\lambda\phi}}, . \quad (3.24)$$

(Brune, 1970) where M_i is the seismic moment (in SI units), $\Theta_{\lambda\phi}$ is the average radiation pattern ($\Theta_{\lambda\phi}$ = 0.55 for S waves (Aki and Richards, 1980; Boore and Boatwright, 1984)), v is the S-wave velocity at the source (we assume on average v =3.5km/s), F is the free surface amplification (F = 2.0 for normally incident SH waves and a good approximation for SV), ξ is a factor equal to 0.67 to account for the partitioning of energy onto the vertical plane (e.g., Newmark and Hall, 1982) and ρ is the average crustal density (ρ = 2800kgm^{-3}). In order to stabilise this inversion we require that we have both greater than 5 observations of each event, and at least 5 observations recorded at each station.

3.5 Data: UK Local Event Dataset

A vertical component velocimetric dataset from 1992 to 2006 was obtained from the British Geological Survey. A number of criteria were applied to the dataset before use. We required that each record must have at least a P- or an S-phase pick; if only one was recorded the other phase time was estimated using a P:S velocity ratio of 1.73. A hypocentral distance of greater than 15km was required to satisfy criteria for the use of point source spectral models (Brune, 1970, 1971; Boatwright, 1978). A 4-second noise window (prior to P-wave arrival or, if unavailable, in the P-wave coda) was extracted along with an 8-second signal window (beginning at the S-wave onset). A signal-to-noise ratio (really the signal plus the noise to noise ratio) greater than 3.0 from at least 2 to 20Hz was then checked for (we found this to be an optimum value, where noise was not influencing the results significantly, whilst maximising the available data). This constraint is applied to each data point in the Fourier spectrum. Any exception to the requirement resulted in the spectrum bandwidth being reduced to remove the data point in question. This prevents erroneous data from influencing the results. If the remaining bandwidth is less than 2 to 20Hz (e.g., 3 to 18 Hz) the record was completely removed. Importantly we required that a seismometer response function was available for the record. This was carefully deconvolved with the signal and noise spectra so that we can set $I_j(f) = 1.0$. We apply an energy integral method commonly used to further refine the signal window and provide a stable estimate of the Fourier velocity spectrum (e.g., Raoof et al., 1999; Atkinson and Boore, 1995, Bay et al., 2003): we define (albeit arbitrarily) that the duration of the S_n or S_g wave is contained within the 5-75% energy integral of the 8 second window for the spectral inversion.

Spectral analysis of the signal and noise windows is performed using the multi-taper method (Lees and Park, 1995; Park et al., 1987) with five 3π–prolate tapers. This method gives vast

improvement over the single taper method giving much smoother spectra, lower spectral leakage and lower variance (e.g., Park et al., 1987). As a result we obtain wider frequency bands above the signal-to-noise ratio, whilst avoiding biasing effects of simply smoothing the spectra. A maximum frequency range of 1 to 25Hz is imposed to avoid frequencies beyond the filter-based maximum frequency of the recorder and below the natural frequency of the seismometer as here the inverse seismometer response function dominates, resulting in an amplification of the instrument noise. Seismometer types used vary considerably in the UK, both spatially and through time, due to a slowly growing network. The majority are single or three-component short-period (1s) seismometers with high-gain recording at 100Hz. Data were also available from a limited number of broadband instruments and accelerometers (after integration of the signal into velocity). However, we use only the vertical component from all instruments. After the selection criteria were applied a total of 3566 records (from 336 events to 115 stations) were available for use in the initial spectral inversion and the tomographic Q inversion. Further data limitations were applied during the tomographic and amplitude inversions resulting in a final count of fully parameterised records of 3230 records of 273 earthquakes obtained at 115 stations. Figure 3.3 shows the distribution of ray-paths and the distribution of both magnitude and source depth against hypocentral distance.

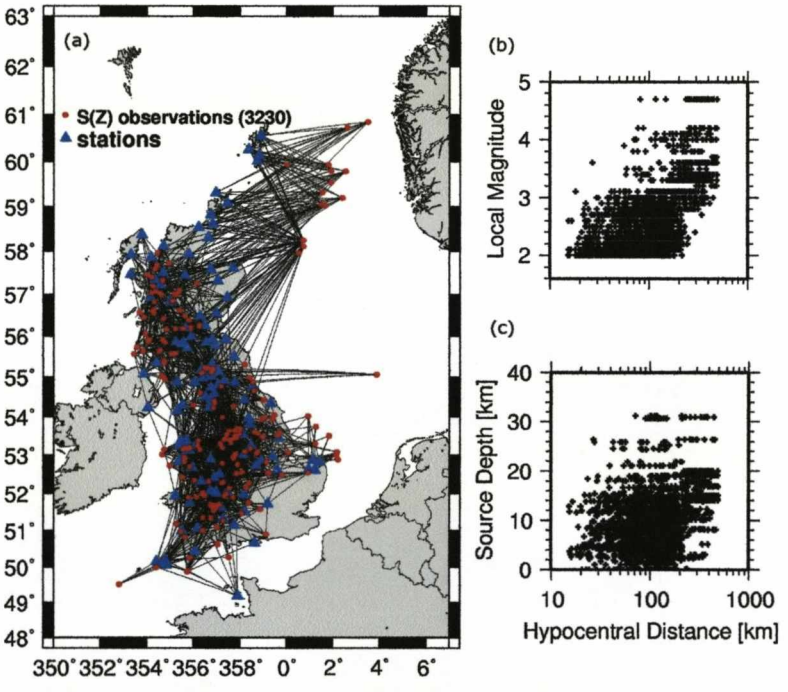

Figure 3.3: a) Plot of ray paths used in all inversions. Circles indicate events, triangles indicate stations and the solid lines joining the two are ray paths. b) Local magnitude – hypocentral distance distribution. c) Source depth – hypocentral distance distribution.

There is some bias for ray-paths to lie to the west of the UK, as this is the area of higher seismicity. Larger magnitude events tend to be recorded at greater distances, as the high

gain instruments do not provide an on-scale recording of high amplitude shaking. The distribution of source depth is fairly uniform down to 15km, becoming sparser at greater depths.

The method described previously is applied to this dataset. We present the results of the log-space L2 fit, which we later find to be the most robust of the minimisation functions.

3.6 Results

3.6.1 Results: Source model

We initially ran the complete inversion procedure using differing source models to find the best fit. Observations of events between $2.0<M_L<3.0$ and between 15km and 25km are used, yielding 33 records. Assuming that little attenuation occurs for these short path lengths we also fix t* = 0, such that the we only test the similarity of the source model to the record. Three source models, all slight deviations from the general form in Equation 3.2, were tested: the Brune (1970; 1971) omega-squared spectrum ($\gamma=2.0, n=1.0$), the omega-cubed spectrum ($\gamma=3.0, n=1.0$), and the Boatwright (1978) spectrum ($\gamma=2.0, n=2.0$). Frequency-independent attenuation is applied. Table 3.1 summarises the overall misfit values using the log-space L2 fit (Equation 3.14). The best fitting source model was found to be the Brune omega-squared spectrum, regardless of the constraint on t*. The Brune omega-squared spectrum was therefore used in the spectral models for this dataset.

Source Model	Log L2 fit	% change	Log L2 fit (t*=0.0)	% change
Brune Omega Squared	0.132	-	0.141	-
Boatwright	0.138	4.55	0.158	12.1
Omega Cubed	0.190	43.9	0.217	53.9

Table 3.1: Selection of source model. The overall fit (sum of the residuals) of the model to the data assuming t is variable and t*=0.0 are shown. The best fitting model is highlighted. The percentage change indicates the increase in the residual fit from the best model.*

3.6.2 Results: Frequency Dependence of Q

The frequency dependence of Q, given by:

$$Q=Q_0 \cdot f^{\alpha} \tag{3.25}$$

is not directly inverted for, instead a grid search around α is performed to find the optimum value for this parameter. We find that without the use of Q tomography, the frequency-dependent $Q=Q_0.f^{0.4}$ relation best fits the shape of the spectra based on the overall residual misfit using the log L2 fit (3.5% better than frequency-independent Q) (Table 3.2).

Frequency dependence (α)	Log L2 fit of spectra	% Change	Log L2 Fit to 1/R spreading	% Change
0.0	0.112199	--	0.43583	--
0.1	0.110938	-1.12	0.452793	3.89
0.2	0.109866	-2.08	0.473619	4.60
0.3	0.108858	-2.98	0.501395	5.86
0.4	0.108263	-3.51	0.540626	7.82
0.5	0.108323	-3.45	0.59895	10.79
0.6	0.109002	-2.85	0.690237	15.24
0.8	0.112934	0.66	1.19757	73.50

Table 3.2: Overall fit residuals from the grid-search of α (the frequency dependence of Q) for both the spectral and amplitude inversions. The percentage change is in relation to the frequency independent Q (α =0.0). Q tomography is not applied.

However, synthetic tests on this dataset showed that it was not possible to accurately resolve α: a mean frequency dependence of 0.3 was found to best fit an input dataset with actual frequency dependence of 0.0. This is due to the effect of $\alpha > 0$, which applies higher attenuation to lower frequencies, and hence, becomes similar in effect to that of geometrical spreading. In fact, for $\alpha = 1.0$, attenuation (note the differentiation between Q, a measure of attenuation and attenuation, the exponential function) becomes completely independent of frequency and the complete frequency range will be reduced by the same factor, dependent on the path travel time, such that we are not able to differentiate between α and apparent geometrical spreading at all:

$$\alpha \to 1.0 \Rightarrow e^{(-\pi \cdot f^{1-\alpha} \cdot t_{ij}^*)} \to e^{(-\pi \cdot t_{ij}^*)} \quad . \tag{3.26}$$

As such, there exists a trade-off between all parameters and the parameter α. Synthetic testing of the dataset (described later) showed that, while the spectra themselves do not allow us to resolve the frequency dependence, if wrong the result can be seen in an increase in the scatter of the signal moments relative to the geometrical spreading model. When we attempted to invert our synthetic dataset (with $\alpha=0.0$) and found that a mean $\alpha=0.3$ best fitted the spectra, a subsequent inversion of the amplitude parameters highlighted this problem. An increase in the misfit to a 1/R geometrical spreading model of 134% was found, compared to the same inversion using a frequency-independent Q. The covariances of all parameters relative to the frequency dependence of Q are determined later in a bootstrap and synthetic analysis (see Table 3.14). From this it is obvious that there is a very large trade-off between the frequency dependence of Q and the signal moment.

In fitting the signal moments from the real data to a geometrical spreading model we find that the best overall fit is found for a frequency-independent Q model, rather than the $Q=Q_0.f^{0.4}$ model (Table 3.2). Taking into account the strong covariance between this frequency dependence, and the signal moment and the increasing scatter of signal moments relative to a geometrical spreading model of 1/R, we proceed assuming frequency-independent Q. Comparatively it seems likely that if we were to initially correct the signal moment using a fixed geometrical spreading model this would result in a strong influence on the observed frequency dependence of Q. Theoretical studies have shown that, at local to regional distances, the rate of geometrical spreading is influenced by sub-surface heterogeneity, such as simple layering (e.g., Campillo et al., 1984). Added to the fact that we observe apparent geometrical spreading, which is dependent on the defined signal window length, it would be convenient to overcome the trade-off between these parameters without fixing the rate of geometrical decay.

3.6.3 Results: Quality Factor

The tomographic inversion requires a known velocity model. We use a model supplied by the British Geological Survey, and given in Table 3.3.

Depth (km)	S Wave Velocity (km/s)
0.0	2.31
2.52	2.31
2.52	3.41
7.55	3.41
7.55	3.73
18.9	3.73
18.9	4.05
34.2	4.05
34.2	4.62
1000.0	4.62

Table 3.3: A general S wave velocity model for the UK. Provided by the British Geological Survey.

We find that a 1D three-layered structure is 52% better in overall spectral fit than the homogeneous Q=1000 model. Within reason, the higher the degree of complexity we allow the Q model to take, the better the fit. However, we found that the 3-layer Q model was the most complex that we could robustly resolve. The number of layers is limited by the depth coverage of the travel paths. A five layer model was initially used, but it proved unstable. The number of ray-paths required to robustly determine the Q value of a node is significantly higher than that required in velocity tomography, which is due to the inherent variance and

covariance of the initial t* estimates used: whereas travel time estimates may be obtained to within ± ~1% of their true value, the error in the t* estimates are orders of magnitude higher. The strong covariance of t* relative to the other parameters being determined also increases the risk of biasing the t* estimates. It is important that we do not simply aim for an increase in fit, regardless of the stability of the solution. We tested the stability of the Q model by using different starting models. The starting models were always homogeneous Q in order to not influence the inversion. Q values of 500, 1000, and 1500 were used, where 1000 was the average Q of the unconstrained spectral inversion. The model was constructed with nodes at 2.5km, 20km and 34km as shown in Table 3.4.

Q_{start}	Initial t* Variance	Q (2.5km)	Q (20km)	Q (34km)	Final t* Variance	% Variance Reduction
500	*3.5e-3*	770	1100	6200	3.2e-4	90.86
1000	*6.9e-4*	990	1000	5500	3.2e-4	53.62
1500	*4.3e-4*	1000	900	*5500*	*3.2e-4*	*25.58*
Average Q		920 ± 130	1000 ± 100	*5700 ± 400*		

Table 3.4: Results of Q at different depths for different starting models. The initial variance from the homogeneous starting model and the reduction in variance using the layered model is shown. The 'average Q' at each depth is the mean and standard deviation of Q from the three starting models. The grey highlighted model is used in this study.

Table 3.4 shows that the inversion was relatively stable, with only the lowermost layer significantly affected by the starting model. However, as Q^{-1} is indicative of attenuation, the large values of Q (~6000) at this depth mean that the layer contributes to t* very little. We adopt the model found using the homogeneous Q = 1000 (Table 3.4) starting model, as this was the harmonic mean of Q from the path variable t* estimates, and provides a consistent basis for a starting model.

3.6.4 Results: Apparent Geometrical Spreading

We use the signal moments to invert for the rate of apparent geometrical decay, moment magnitudes and site correction terms. The rate of apparent geometrical spreading, mainly due to geometrical spreading, was inverted for using a tri-segment structure (Equation 3.5). The starting model is obtained from inspection of Figure 3.2 as being segmented at distances of 50km and 150km. We choose this model because at 50km the path dependent duration of the S wave seems to stop growing; then at 150km the L_g wave begins to disperse from the S wave (Figure 3.2). We then allow the grid-search to go from 35 to 70km in steps of 5km for the first segmentation and from 80km to 200km for the second segmentation in steps of 10km. A decision was made to fix the exponent of the first segment to 1.0. This was required to stabilise the inversions further, and prevent any influence of the Q model on the

geometrical spreading model.

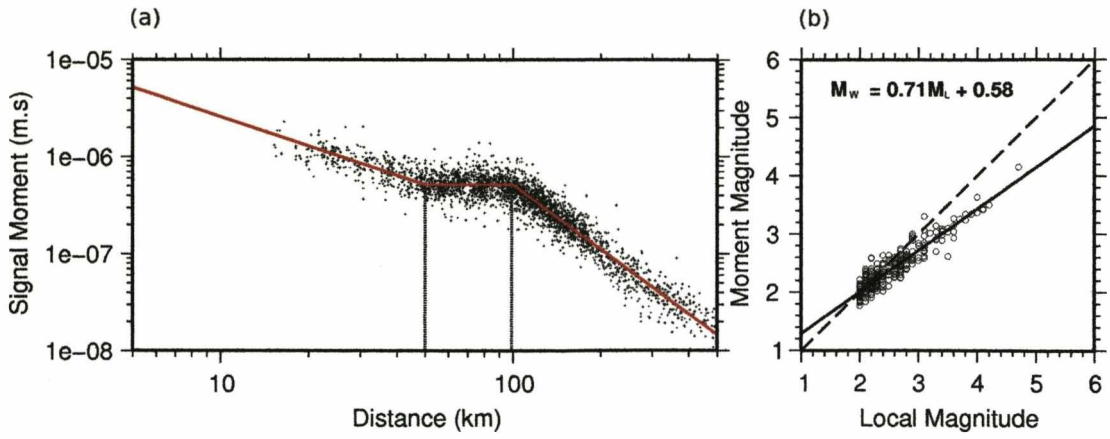

Figure 3.4: (a) Plot of signal moments normalised to a magnitude M_w = 3.0 event and station correction of 1.0. The solid trend line is the apparent geometrical spreading function. 'Distance' refers to hypocentral distance. (b) Plot of recorded local magnitude against computed moment magnitudes. The dashed line indicates M_L=M_w, the solid line shows the L2 trend of data (Equation 3.32).

We find a multiple segment model of $R_0/R^{1.00}$ (fixed) for hypocentral distances of up to 50km, roughly the crustal thickness; followed by $R_1/R^{0.00}$ up to 100km; then $R_2/R^{2.22}$ for distances beyond 100km (Figure 3.4). The rate of decay used in the local magnitude scale calibrated for California (Hutton and Boore, 1987), and used to determine the local magnitudes of events in the UK is:

$$M_L = \log(A) + 1.11 \log(R) + 0.00189\, R - 2.09 \qquad (3.27)$$

where A is the maximum ground amplitude (zero to peak in micrometers) and R is the hypocentral distance in km. This is a slightly faster rate of decay than we used for the UK, and may explain why we see a gradient between M_L and M_w not equal to unity. However, changing the fixed exponent from 1.0 results in a constant shift in the moments and so does not rectify the 0.7 gradient found. The complexity of the geometrical spreading model lies in the fact that beyond distances roughly equal to the crustal thickness the crust begins to act as a wave-guide. Initially we see reflections, for instance from the Moho, then beyond 100km surface waves begin to develop. The first segmentation (at 50km) may be interpreted as the transition from S_g only to a combination of S-type phases (S_g, sS, S_n, etc.) at distances greater than 50km. The second segmentation of the geometrical spreading function leads to a super-spherical decay. This may be due to the domination of the signal window by the S-wave coda at distances beyond 100km. Rates of coda decay have been quoted at up to $1/R^2$ (e.g., Frankel, 1991), which is consistent with the final segment of the geometrical spreading curve.

3.6.5 Results: Moment Magnitudes and Stress Drops

A result of our final inversion is the long-period plateau value at the earthquake source. Using Equation 3.24 to compute the seismic moments, we can then compute the moment magnitudes of the earthquakes:

$$M_w = \frac{2}{3}\log M_0 - 6.03 \ , \tag{3.28}$$

where the seismic moment, M_0, is in SI units. Combining the seismic moments with the source corner frequencies we are able to compute the Brune stress parameter. The theoretical stress drop of a circular rupture is given by:

$$\Delta\sigma = \frac{7}{16}\cdot\frac{M_0}{r_0^3} \ , \tag{3.29}$$

(Eshelby, 1957) where the radius of the rupture, r_0, is given in Equation 3.6. Using the apparent geometrical spreading model above we find the computed stress drops are of the order of 0.1 to 10 MPa (Figure 3.5a).

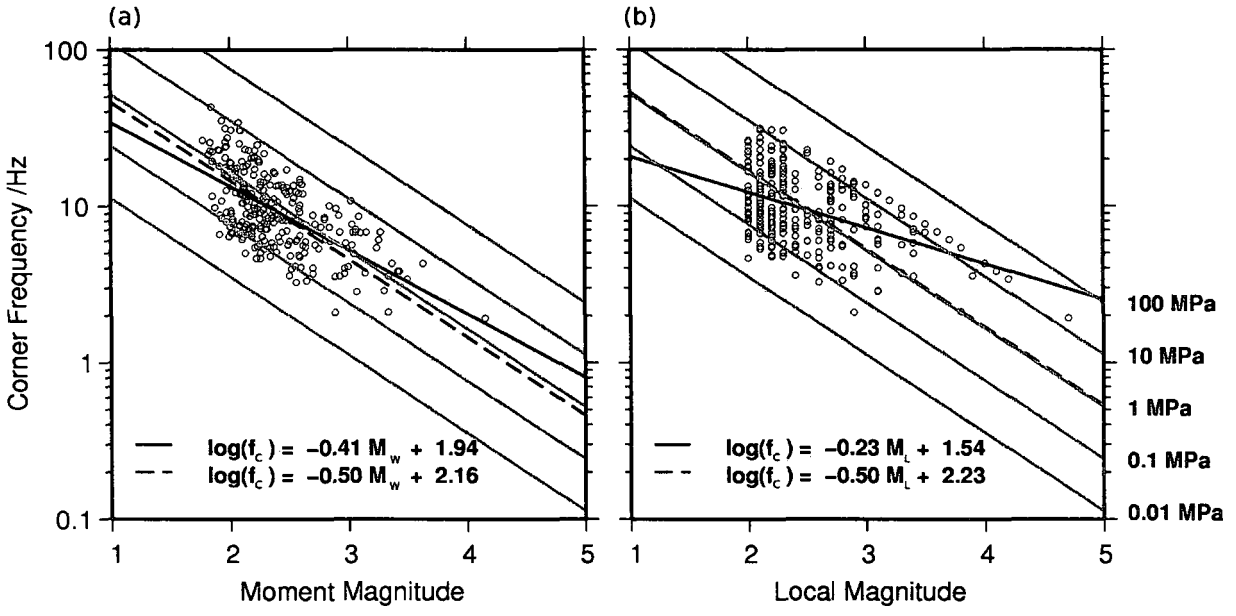

Figure 3.5: Plot of source corner frequency against a) computed moment magnitude and, b) assumption of M_w=M_L for the same events. Grey lines indicate constant stress drop for a circular rupture. The solid black line indicates L2 trend of data whilst the dashed line indicates constant stress drop trend.

Assuming that the magnitudes are known (which is reasonable as the error in the moment is much less than the error in the source corner frequency), a least-squares best fit relationship of:

$$log(f_c) = -0.41M_w + 1.94, \tag{3.30}$$

is found between the source corner frequencies and moment magnitudes. This equates to a small increase in the average stress drop with increasing magnitude. The relationship is poorly determined, however, with a coefficient of determination (R^2) of only 0.35. This is due to the cluster of low magnitude events and relatively few higher magnitude (M_w>3) events along with a broad range of stress drops even within this small magnitude range. Forcing a constant average stress drop we obtain a relationship of:

$$log(f_c) = -0.50M_w + 2.16, \tag{3.31}$$

which gives a mean of 0.68MPa with a log-normal standard deviation of 0.58 such that the 68.3% confidence interval is between 0.18 and 2.6MPa. A plot of the spatial distribution of stress drops is shown in Figure 3.6.

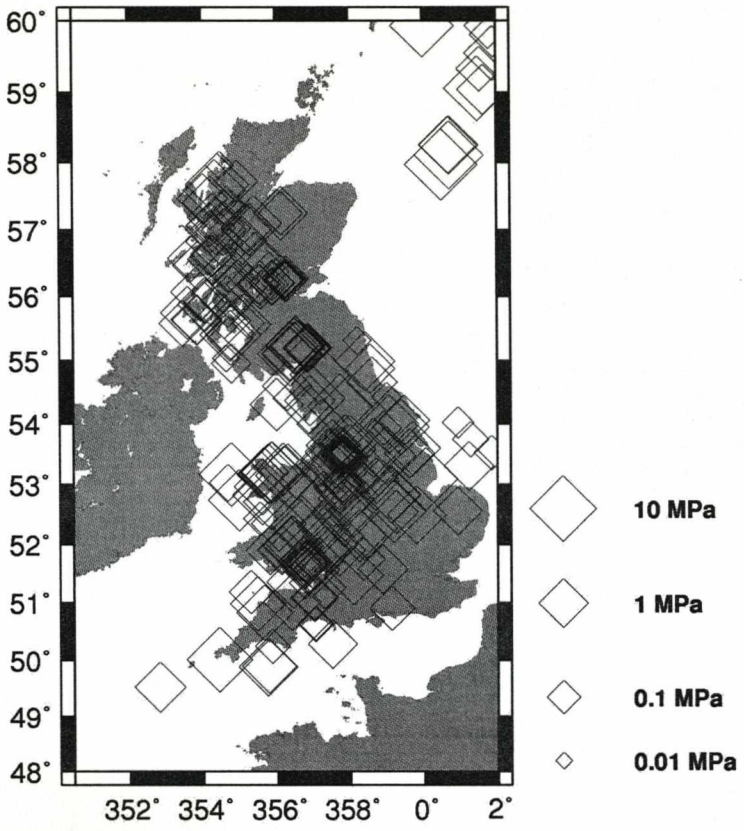

Figure 3.6: Plot of spatial distribution of stress drop parameters. The magnitude of the stress drop is indicated by the size of the symbol, as shown in the key.

This plot highlights a possible reason for a relatively large range of stress drops over the whole of the UK. There exists clustering of earthquakes with stress drops of a similar order of magnitude in very localised regions. Considering these limitations Equation 3.30 should not be taken as an equation into which we can plug any magnitude. If we consider that the stress drop is not magnitude dependent then Equation 3.31 should be used and may prove to be more reliable at higher magnitudes, but we must also consider the variance of the stress drop when using these equations.

A relationship (R^2 = 0.87, correlation coefficient $\rho_{Mw,Ml}$ = 0.93) between recorded local magnitude, M_L, and moment magnitude is found as:

$$M_w = 0.71M_L + 0.58, \tag{3.32}$$

(Figure 3.4b). If we choose to ignore Equation 3.32 and set $M_w = M_L$; as required for earthquake self-similarity (Deichmann, 2006), then we find that the variance of stress-drop parameters increases, and that for higher magnitude earthquakes we have higher stress parameters. The theoretical relationship of $M_L=M_w$, subject to conditions of self-similarity (see

Deichmann, 2006), does not apply to this magnitude range. This would suggest that, on a regional scale, there is no self-similarity of earthquakes across the magnitude range $2 < M_L < 4.7$. Another reason for not finding a $M_L=M_w$ relationship may lie with the use of an incorrect local magnitude scale: the M_L scale assumes that the record has lost much more energy than it actually has, and hence, the source magnitude is assumed to be higher to compensate. Consequently the local magnitudes are on average higher than the moment magnitudes, which are a direct measurement of radiated energy. The increasing average recording distance with increasing magnitude (Figure 3.3b) results in larger magnitude events being affected to a greater degree by this difference. The bootstrap analysis, described later, resulted in many individual computations of each moment magnitude using different data. The average standard deviation of the moment magnitude was around $\pm 1.7\%$.

3.6.6 Results: Site Response Functions

Site response functions represent the station-specific effect on the record. They are inverted from both the amplitude- and frequency-dependent inversions. We obtain a constant amplitude correction representing the mean amplification at the site over all frequencies with respect to the network average (Figure 3.7a) due to, for example, regional level focusing of energy along wave-guides.

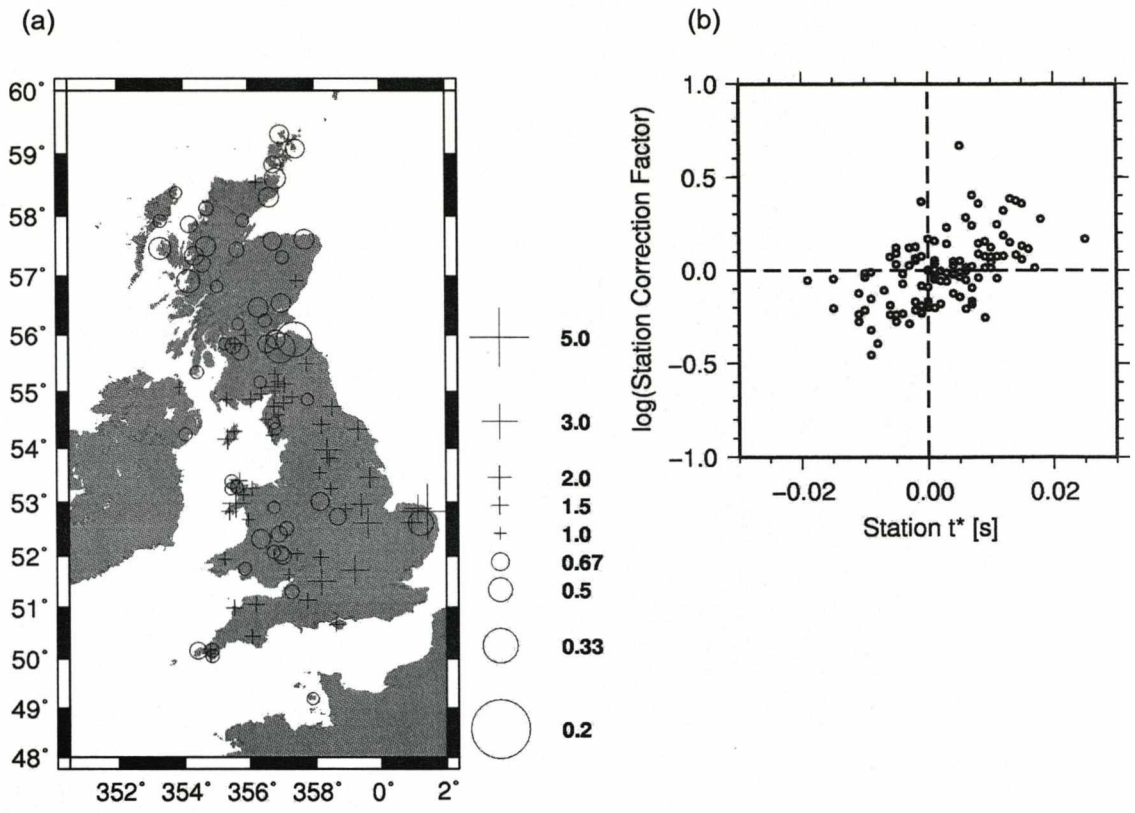

Figure 3.7: (a) Plot of spatial distribution of station amplitude correction factors. Circles indicate de-amplification, crosses indicate amplification. (b) Relationship between station amplitude correction and station t (κ_j) value.*

We find that there is a distinct distribution of the mean site corrections (Figure 3.7b), which correlates to geology on a regional scale. The softer Cenozoic rocks in South East England tend to amplify the signal relative to Wales and Scotland which are predominantly composed of harder Palaeozoic or older rocks. Sites in close proximity to one another tend to share amplitude and direction of amplification; as such we can see clustering in the distribution of the site corrections. Although not fully understood, the distribution of these station-dependent mean amplification factors suggests that they are indeed representing the focusing and defocusing of energy at local to regional levels.

In addition we obtain a frequency dependent response function (e.g., Figure 3.8), due to resonant frequencies caused by near-surface fractured and weathered layers, along with the κ_j at the site (Figure 3.9). The overall response function is a multiplication of these terms in the frequency domain (Equation 3.7).

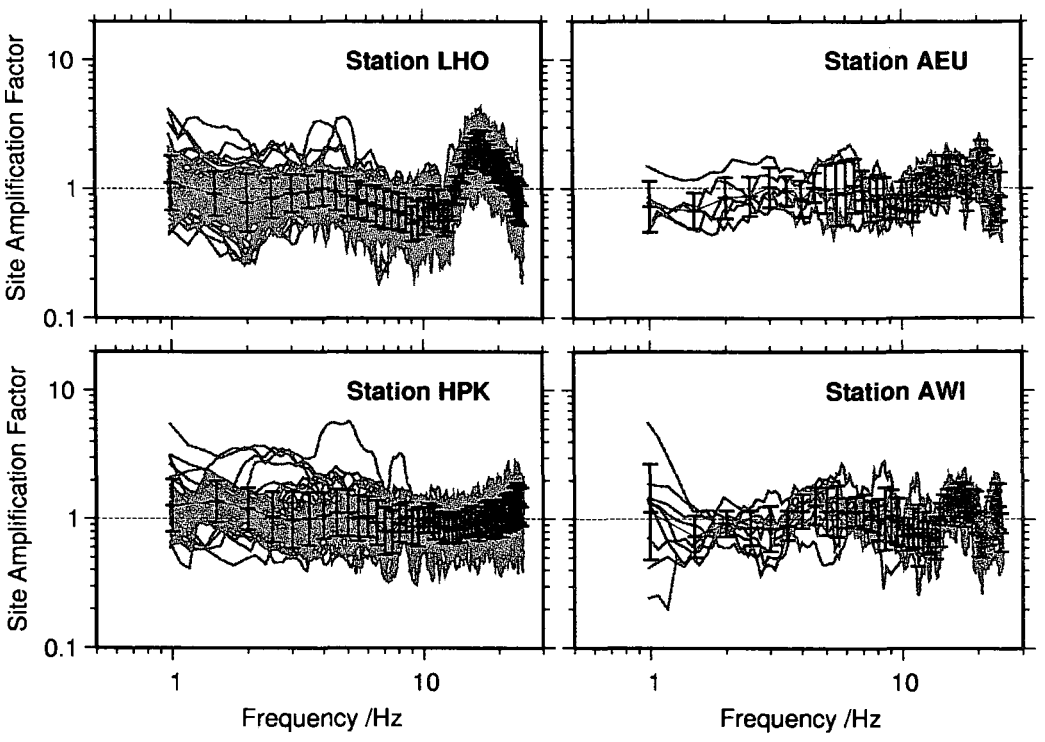

Figure 3.8: Plot of frequency dependent site response for stations LHO, HPK, AWI and AEU. Grey lines indicate individual response functions, black line indicates geometric mean site response function with error bars of one standard deviation.

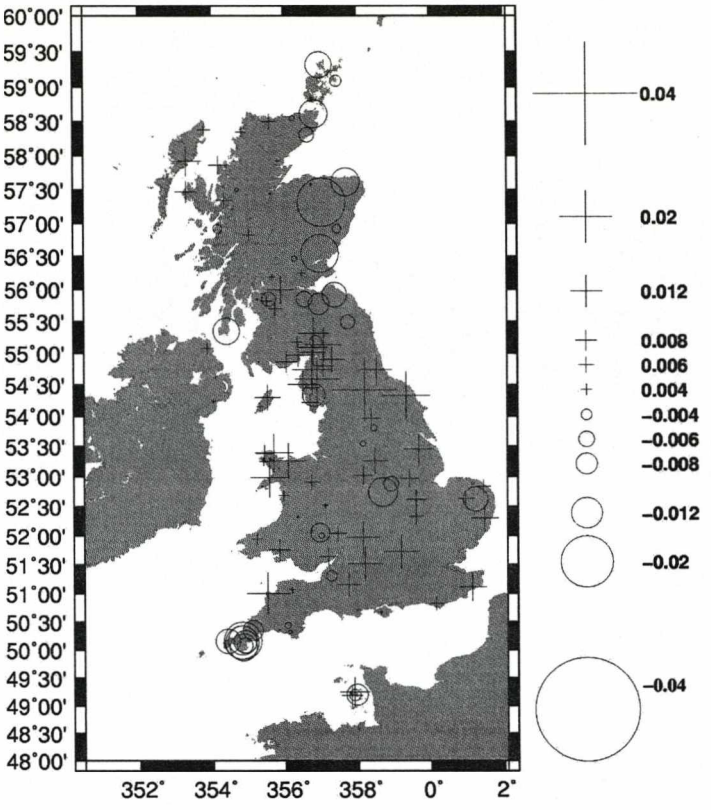

Figure 3.9: Plot of spatial distribution of station attenuation parameters. Circles indicate 'negative attenuation', crosses indicate normal attenuation. Units of κ_j are in seconds.

A comparison of site amplification corrections and site attenuation (κ_j) parameters shows a positive correlation (Figure 3.7b). This implies that stronger site attenuation is found with a stronger site amplification correction: on a regional scale we therefore find that softer, more porous rock regions act to both amplify the whole frequency range, while attenuating the high frequencies more strongly. If we neglect to include the mean site amplification factors we see significantly larger residuals (the average residual being around 30% higher). The detailed interpretation of individual site responses is beyond the scope of this paper; however, the correlation of site attenuation and site amplification could easily be explained by considering the possible causes of mean amplification (a soft rock site), and the causes of strong local attenuation (partially saturated porous soil). These conditions are themselves correlated, in that they generally occur together. Thus we may have a hard rock site, with little soil coverage, leading to mean de-amplification and low site-attenuation; or a soft rock site with thick soils, resulting in relative amplification, along with strong site attenuation.

The residuals from the complete inversion are plotted in Figure 3.10. The residuals show model/data, where the model does not include the frequency-dependent site response ($A_i(f)$).

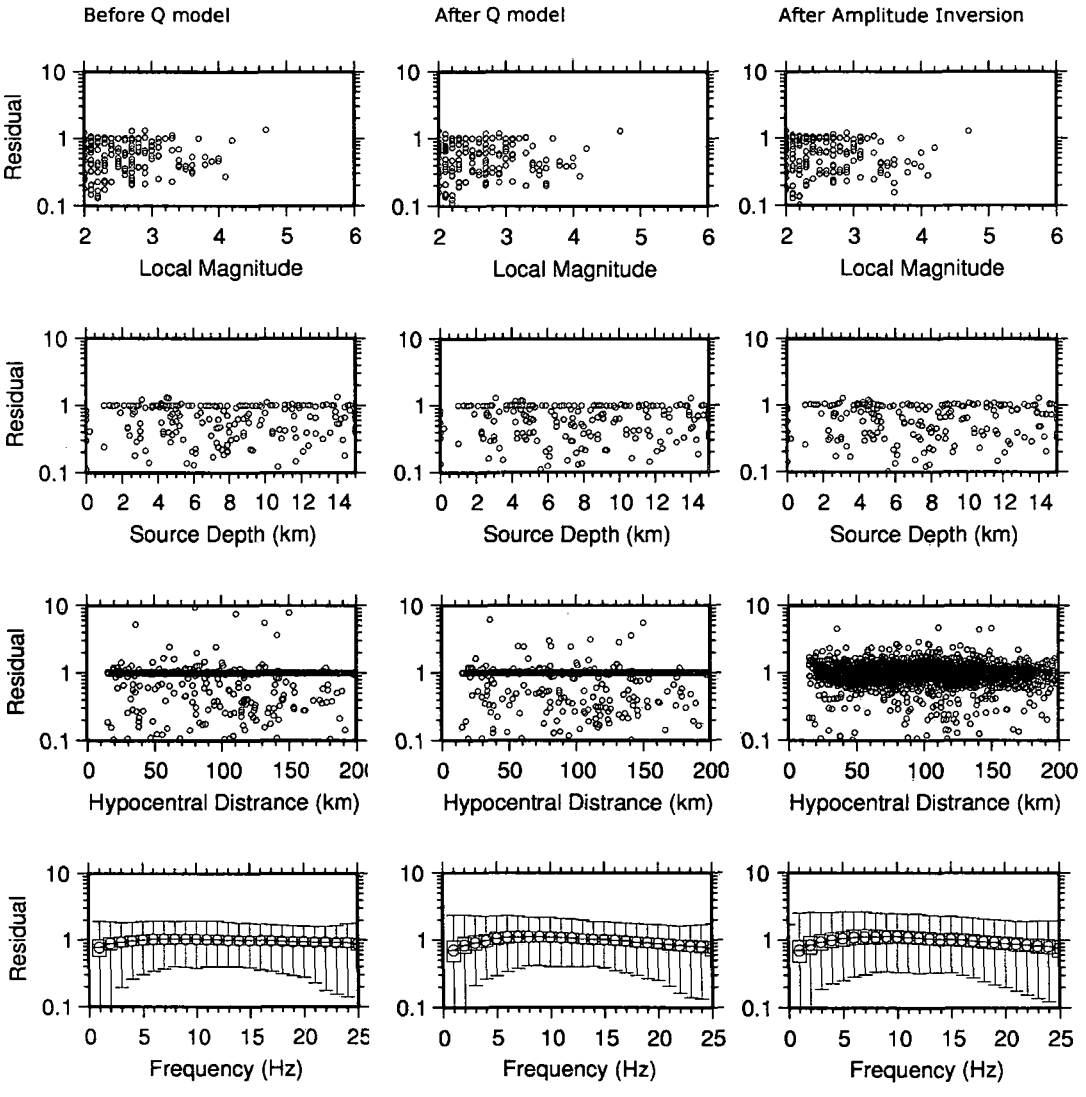

Figure 3.10: Overall residuals (model/data) plotted against frequency, hypocentral distance, source depth and local magnitude. In the frequency plot the error bars show the 68.3 and 95% confidence intervals as the box and whisker respectively. The other plots show the log-mean residual at each distance, magnitude or depth available. The residuals are plotted for the initial inversion (for estimating the t parameter) on the left, the final spectral residuals (after fixing the Q structure) are plotted in the middle, and the final residuals (after the amplitude parameter inversion) are plotted on the right.*

$A_j(f)$ itself is constructed from residuals and may therefore mask any dependencies introduced into the overall residuals. The residuals are plotted both for the initial estimates of the spectral model, for the Q model constrained results and finally after the amplitude parameter inversion. We see no dependency of the residuals on frequency, hypocentral distance, depth or magnitude for either inversion.

3.7 Inversion Robustness and Error Estimation

3.7.1 Synthetic Data

A synthetic database is created by the summation of an amplitude spectrum (computed by forward modelling the results of the full inversion) and the original observation's noise spectrum. We then apply a frequency-dependent site response. This response was determined using the results from the log L2 fit of the real data (e.g., Figure 3.8). The station amplifications are defined by a mean and normal distribution in the log space. A random selection from this distribution at discrete frequencies is used to construct the response with logarithmic interpolation between those frequencies. We invert this synthetic dataset using the initial parameter estimation method as used for the real data (i.e., the stage prior to Q tomography). Table 3.5 shows the covariance matrices of the percentage error between the input and inverted values of the three spectral parameters, the signal moment, the t* value, and the source corner-frequency using the L1 (Equation 3.10), L2 (Equation 3.11), Maximum-likelihood (Equation 3.12), log-space L1 (Equation 3.13) and log-space L2 (Equation 3.14) optimisation functions. This format of covariance matrix table (e.g., Table 3.5) is frequently used throughout the paper, hence a brief description is necessary: the inverted spectral parameters head the columns and rows (in this case a super-heading indicates the inversion type used, but the general form is the same), the cell in column headed 'a' and row headed 'a' shows the variance of parameter 'a'; the cell in column headed 'a' and row headed 'b' shows the value of the covariance of parameter 'a' and 'b', etc.

	L1			L2			ML			Log L1			Log L2		
$\sigma_i\sigma_j$	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
a	0.0124	--	--	0.0197	--	--	0.0124	--	--	0.00896	--	--	0.00861	--	--
b	0.00739	0.598	--	0.00049	0.930	--	-0.0045	0.615	--	-0.0002	0.472	--	-0.0043	0.459	--
c	-0.0118	0.144	0.206	-0.0064	0.193	0.220	-0.0125	0.177	0.239	-0.0129	0.117	0.170	-0.0126	0.126	0.182
Mean (%)	120.04	63.00	132.39	127.79	66.24	145.44	133.45	65.15	137.10	123.90	54.74	104.25	126.09	58.19	108.90

Table 3.5: Synthetic test analysis of the log(parameter percentage deviation) variance, covariance and mean percentage error. 'a' is log($\Delta\Omega_{ij}(r)$ (%)); 'b' is log(Δt^*(%)); and 'c' is log(Δf_c(%)). The first row shows the minimisation function used. The inverted spectral parameters then head the columns and rows, the cell in column headed 'a' and row headed 'a' shows the variance of parameter 'a'; the cell in column headed 'a' and row headed 'b' shows the value of the covariance of parameter 'a' and 'b', etc. The lowermost row is reserved for the mean (percentage) deviation. 100% indicates that the recovered value is equal to the input value. Boxes highlighted dark grey indicate the minimum covariance for that particular parameter pair; boxes highlighted light grey indicate the second best (co)variance. No Q tomography is used – variable t.*

The covariance matrices in Table 3.5 show a high covariance between t* and f_c when we do not use Q tomography for all minimisation functions, highlighting the strong trade-off between these parameters. There is also some trade-off between the t* parameter and the signal moment. The lowest variance for the signal moment and t* is obtained using the log-space L2 fit; for the source corner frequency the log-space L1 fit is better. In Table 3.5 we highlight the best variance or covariances for each parameter pair. Overall it is clear that the log-space fits are superior for this application.

Covariance	log(Δ $\hat{\Omega}_{ij}$ (%))	log(Δt*(%))	log(Δf_c(%))
log(Δ $\hat{\Omega}_{ij}$ (%))	0.013718 (0.008613)	--	--
log(Δt*(%))	0.001521 (-0.004262)	0.010823 (0.458865)	--
log(Δf_c(%))	-0.007042 (-0.012634)	0.001019 (0.126405)	0.005990 (0.181942)
Mean % Deviations	118.31 (126.09)	78.26 (58.2)	76.94 (108.90)

Table 3.6: Variance, covariance and mean of the log(percentage deviation) of results from the log-space L2 inversion of the synthetic data. The inverted spectral parameters head the columns and rows, the cell in column headed 'a' and row headed 'a' shows the variance of parameter 'a'; the cell in column headed 'a' and row headed 'b' shows the value of the covariance of parameter 'a' and 'b', etc. Q tomography constrained results precede bracketed unconstrained results. 100% indicates that the recovered value is equal to the input value.

We now proceed with the inversion, using only the log-space L2 minimisation. Table 3.6 shows the covariance matrices for the log-space L2 fit both before (as with Table 3.5) and after the application of Q tomography. After applying the Q tomography method we see a significant decrease in all covariances apart from the signal moment variance. On average the synthetics show that we will over-predict the signal moment by around 20%, whereas we will under-predict both the source corner frequency and the t* value by around 20%. The signal moment shows stronger variance in deviation from its true value after fixing t*; this is due to the effect of fixing t* and the covariance between t* and the signal moment. Overall it is clearly beneficial to apply Q tomography to stabilise the inversion and, provided that we adopt a sensible Q model, will reduce the errors in our inverted parameters. The small increase in the signal moment parameter, and the covariance between it and t*, mean that a further constraint is required to prevent trade-off between the geometrical spreading function, the Q model and the seismic moments. We achieve this by forcing the exponent of the first

segment of the apparent geometrical spreading model to equal 1.0. In doing this we later find that the computed moment magnitudes vary by less than 2% using the bootstrap method.

In considering a sensible model for the Q structure we must balance the complexity and robustness of the inversion. More complexity (the most 'complex' being path dependent t* as with the initial inversion) in the model leads to poor resolution of the spectral parameters, as highlighted in Table 3.6. In this case a three-layer model was chosen, as a five-layer model proved unstable given different starting Q. As we constructed the synthetic dataset using our inverted Q model, we are able to test the inversion in terms of its ability to correctly reconstruct the Q model.

Depth:	2.5km	20km	34km
Input Q	990	1040	5500
Output Q	1100	1200	5600

Table 3.7: Input Q model and inverted output Q model from the inversion of synthetic data.

Table 3.7 shows the input and inverted Q models. The Q models are very similar: the increasing Q with depth is clearly seen, with a small increase from 2.5 to 20 km followed by a large increase to 34 km. A constant offset of ~100 is found, which is consistent with the 20% lower average t* values obtained from the initial inversion (Table 3.6).

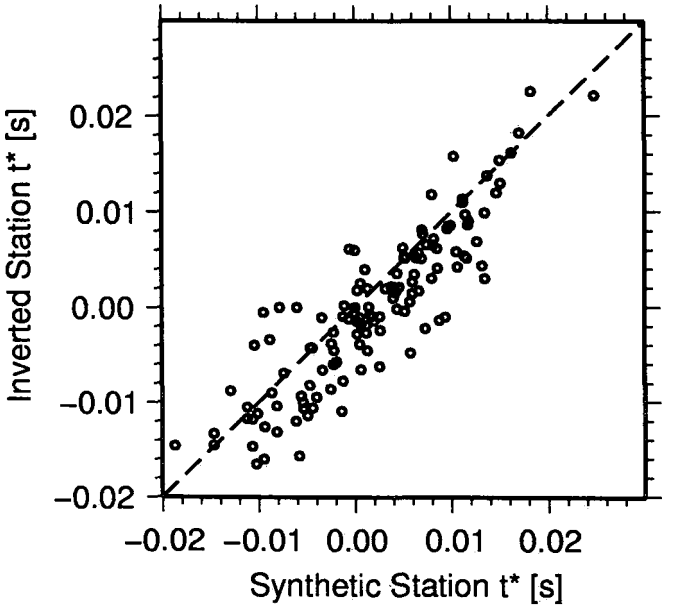

Figure 3.11: Plot of input (synthetic) against inverted station t (κ_j) for the synthetic dataset.*

Figure 3.11 shows the input and inverted κ_j values using the synthetic data. The average input κ_j was 0.0022 with a standard deviation of 0.0081. The inverted mean κ_j was 0.0000 (the zero value having no obvious implication), with standard deviation 0.0083. The average error was -45%. The average negative shift in κ_j values may be attributed to the mean underestimation of the initial t* values (remember that the initial t* estimates are combined path t* and site kappa), as with the Q model. Plotting input against inverted κ_j (Figure 3.11) we find that the general form of the κ_j distribution is retained. As a result we may interpret any distribution of κ_j, but should avoid interpreting the specific amplitude of individual station κ_j values.

3.7.2 Bootstrap Data

Further to the synthetic tests we obtain error estimates by means of a bootstrap analysis. Selection, with replacement of 20% of the data, from the full database is used to create around 160 random sub-sets of the data comprising of 80% non-repeatable and 20% repeatable data. This is done for each optimisation function. Analysis of the statistical deviation of the newly inverted parameters to those values obtained with the full database is then possible. The random selection with repetition of events for the data subsets means that these statistical distributions take into account any data bias such as the inability to model certain events using our standard inversion model. The confidence intervals computed for each parameter therefore indicate to what extent the results are dependent on the data selection. This is performed for both Q tomography constrained values and, for reference, results without Q tomography (Table 3.8 shows the complete distribution results for the log-space L2 fit).

Q tomography	Parameter		<- Stability of deviation statistics ->			
			-σ	Median	+σ	(+σ)- (-σ) *
No	t^*	-σ	-87.3	-85.4	-81.8	5.5
		Median	-17.4	-14.5	-10.7	6.7
		+σ	63.5	66.8	71.9	8.4
		(+σ)- (-σ) **	150.8	152.2	153.7	
Yes	t^*	-σ	-9.9	-5.8	-1.3	8.6
		Median	1.8	5.1	8.8	7
		+σ	13.9	17.9	23.6	9.7
		(+σ)- (-σ) **	23.8	23.7	24.9	
No	f_c	-σ	-47.5	-46.5	-44.3	3.2
		Median	-21.4	-17.5	-14.1	7.3
		+σ	307.9	362.6	389.1	81.2
		(+σ)- (-σ) **	355.4	409.1	433.4	
Yes	f_c	-σ	-9.3	-4.8	0.0	9.3
		Median	0.0	5.0	10.2	10.2
		+σ	15.8	21.6	27.6	11.8
		(+σ)- (-σ) **	25.1	26.4	27.6	
No	Signal moment	-σ	-21.5	-20.7	-20.1	1.4
		Median	11.5	12.6	13.8	2.3
		+σ	64.5	66.3	68.5	4
		(+σ)- (-σ) **	86	87	88.6	
Yes	Signal moment	-σ	-14.6	-13.6	-12.5	2.1
		Median	-1.2	-0.4	0.5	1.7
		+σ	13.7	15.1	16.6	2.9
		(+σ)- (-σ) **	28.3	28.7	29.1	

*Table 3.8: Bootstrap analysis results showing parameter variation based on data selection with replacement using the log-space L2 fit. The values are the percentage deviation from the results found using the complete dataset (0% indicates that the recovered value is equal to the reference value). Each parameter is represented twice; before, and after Q tomography. In each case the median, and lower and upper 68.3% confidence limits of the inverted parameter are shown (reading up-down the table). For each of these statistics we also show its stability (reading left-right). So the median of the median parameter deviation lies in the centre of each 3 by 3 square of values. * is a measure of the parameter statistic stability; ** is a measure of the inversion robustness (lower is better in both cases).*

An advantage of the bootstrap method is that the stability of the error estimates may be obtained (these are displayed reading left to right in Table 3.8; so for each parameter statistic, for example the median, we read left to right to find lower, median, and upper confidence limits of that statistic). As such we not only obtain error (or deviation from our norm) estimates, but we may also assess their validity. We find the stability of the parameter deviation statistics is very good, with only very small changes depending on the sub-dataset. We find significant improvements in the robustness of results (now reading up-down in Table 3.8) and therefore the errors associated with those results by using the Q tomography method. The bootstrap analysis allows us to observe the robustness of results from different fitting functions. This is displayed in Table 3.9 in terms of the range between the median 68.3% confidence intervals for each of the three parameters after Q tomography is applied.

Fit type	$\hat{\Omega}_{ij}(r)$	t^*	f_c
Log L2	29	24	26
Log L1	35	26	24
L1	52	40	76
L2	88	80	166
Maximum-Likelihood	64	80	50

Table 3.9: Measure of inversion robustness by optimisation type and parameter. Values represent the range between the median-upper and median-lower 68.3% confidence values (i.e., the numbers highlighted bold in Table 3.8) in terms of percentage deviation from the original results using the log-space L2 fit.

Table 3.9 shows that the log-space fits are the most robust. If we assume that the parameter deviations in the bootstrap analysis (Table 3.8) are log-normally distributed we may summarise the data in Table 3.8 using the covariance matrix. Table 3.10 shows the covariance matrices derived assuming log-normally distributed percentage error for the stages prior to, and subsequent to, Q tomography using the log-space L2 fit.

Covariance	log(Δ $\hat{\Omega}_{ij}$ (%))	log(Δt*(%))	log(Δf$_c$(%))
log(Δ $\hat{\Omega}_{ij}$ (%))	0.005354 (0.027654)	--	--
log(Δt*(%))	0.001909 (0.022028)	0.004802 (0.674385)	--
log(Δf$_c$(%))	-0.002866 (-0.007421)	0.001355 (0.136515)	0.005112 (0.172602)
Mean % Deviations	99.88 (114.87)	106.00 (64.59)	106.68 (130.51)

Table 3.10: Results of the bootstrap analysis showing variance, covariance and mean deviation of the log(percentage deviation) of results (100% indicates that the recovered value is equal to the input value) using the log-space L2 inversion assuming a log-normal distribution. The inverted spectral parameters head the columns and rows, the cell in column headed 'a' and row headed 'a' shows the variance of parameter 'a'; the cell in column headed 'a' and row headed 'b' shows the value of the covariance of parameter 'a' and 'b', etc. Q tomography constrained results precede bracketed unconstrained results.

Prior to applying Q tomography we observe significantly higher variance in the corner frequency and t* parameters. As with the synthetic testing we also see the high covariance between the signal moment and t*; and the source corner frequency and t*. After constraining the Q model we significantly reduce all covariances and variances with exception of the signal moment which suffers due to the covariance with the fixed t* value. The variances equate to the following parameter variability: for log(signal moment) $\pm 6.9 \times 10^{-2}$; for log(t*) $\pm$ 7.2×10^{-2};and for log(f_c) $\pm$ 7.3×10^{-2}. Before we constrained Q the parameter variabilities were approximately: $\pm$ 1.7×10^{-1} for the log(signal moment); $\pm 8.0 \times 10^{-1}$ for log(t*) and $\pm$ 4.2×10^{-1} for log(f_c).

Tables 3.8 and 3.10 also show that the mean deviations after Q tomography were low and as such our results using the full dataset lie in the centre of the likelihood distribution. Comparing Tables 3.10 and 3.6, the covariance matrices using bootstrap data and synthetic data are somewhat similar. This suggests that the synthetic dataset has been appropriately constructed, and that the bias of results (i.e., the mean deviation) found in Table 3.6 is correct.

Unlike for the spectral parameters, the stability of the error estimates for the geometrical spreading exponent could not be ascertained due to the time that this would take, as we only compute three exponents per dataset, in comparison to, for example, hundreds of source corner frequencies. The deviation of the geometrical spreading exponent (Table 3.11),

dependent on the data selection, shows no significant reduction in the confidence interval, depending on whether or not we use the Q tomography: we only see a baseline shift of all statistics.

Q tomography	*Parameter*		*% Deviation from final results*
	λ_1		*(fixed)*
No	λ_2	$-\sigma$	-0.4
		Median	18.2
		$+\sigma$	35.2
Yes	λ_2	$-\sigma$	-40.6
		Median	-22.9
		$+\sigma$	-2.0
No	λ_3	$-\sigma$	-25.4
		Median	-21.7
		$+\sigma$	-16.7
Yes	λ_3	$-\sigma$	2.9
		Median	6.6
		$+\sigma$	10.5

Table 3.11: Bootstrap analysis of the error distribution of geometrical spreading decay values using the log-space L2 fit. Results are shown before and after Q tomography is applied. The values are the percentage deviation from the results found using the complete dataset.

Without using Q tomography, the first non-fixed geometrical spreading exponent (λ_2) is around 30-40% higher. We also see that the full dataset geometrical spreading exponents (which have deviation of 0.0% by definition) lie outside the 68.3% confidence interval defined by this range. The covariance matrices (assuming a normal distribution, and the correct separation distances - which are the same as used for the full UK dataset) of the geometrical decay were computed and are shown in Table 3.12. Table 3.12 gives standard deviations for the geometrical spreading exponents as ± 0.06 for λ_2 and ± 0.16 for λ_3.

Covariance	λ_1	λ_2	λ_3
λ_1	0.0 (fixed) (0.0)	---	---
λ_2	0.0 (0.0)	0.00302 (0.002286)	---
λ_3	0.0 (0.0)	-0.00384 (-0.007043)	0.0244 (0.050038)

Table 3.12: Covariance matrix of geometrical spreading exponents in the tri-segment model from the bootstrap analysis. The inverted spectral parameters head the columns and rows, the cell in column headed 'a' and row headed 'a' shows the variance of parameter 'a'; the cell in column headed 'a' and row headed 'b' shows the value of the covariance of parameter 'a' and 'b', etc. Values in this case are not percentage deviations, but absolute deviations. Covariances in brackets are before Q tomography has been applied, those not in brackets are after Q tomography has been applied.

Error Estimate	Bootstrap A	Bootstrap B	Synthetics
Parameter			
$\hat{\Omega}_{ij}$	±14%	±17%	±32%
t*	±12%	±17%	±19%
f_c	±13%	±18%	±14%

Table 3.13: Estimated error of the inverted parameters. 'Values are approximated by taking half the range between the upper and lower 68.3% confidence intervals. 'Bootstrap B' and 'Synthetics' estimates are assuming log-normally distributed percentage deviations.

Estimates of the error in our inverted spectral parameters are given in Table 3.13 using the different distributions and the synthetic and bootstrap data. Based on all the results we chose to use the log-space L2 optimisation throughout this paper. In Table 3.13 we see that the computed variances assuming log-normal distributions (Tables 3.6, 3.10 and 3.12) relate well to the results in Tables 3.8 and 3.11 which use the 68.27% confidence interval and do not assume any distribution for t* and f_c. Both the bootstrap and synthetic analyses show that using the tomographic Q constraint gives a significant reduction in the range of errors for the 68.27% confidence interval. A benefit of using the synthetic testing is that we can observe any systematic bias in the inversion: we see that the mean error is an overestimation of the signal moment and an underestimation of t* and f_c, while the bootstrap is considered a more robust determination of variance. The correlation of results from these independent methods indicates that we can rely on the bias and covariance statistics from the synthetic tests, and that the successful reconstruction of the Q model means that we are not simply reducing the variance of results by adopting an inaccurate Q model in the bootstrap analysis. The covariance of t* and the signal moment, which causes an increase in the signal moment

variance after fixing t*, does not seem to have a significant effect on the computation of the moment magnitude, with variances of less than 2% propagated into the moment magnitudes. Instead we see a change in the geometrical spreading model, implying a strong covariance between the Q model and the geometrical spreading model. Table 3.14 shows the covariance between the frequency dependence of Q (α) and the parameters from the spectral inversion.

Covariance	$\log(\Delta \hat{\Omega}_{ij}$ (%))	log(Δt* (%))	log(Δf$_c$ (%))
α (absolute value)	0.785	0.0468	0.118

Table 3.14: Covariance parameters relating to the absolute value of α (the frequency dependence of Q) and the resulting log(percentage deviations) of the other spectral parameters prior to Q tomography.

3.8 Discussion and Conclusions

Significant trade-offs exist in the fitting of Fourier spectra. It is insufficient to only constrain the corner frequency as very high errors result. However, the use of Q tomography takes a significant step in addressing this problem. We showed this by reconstructing and performing the inversions on a synthetic version of the database and performing a bootstrap analysis on the original dataset. Errors were shown to be significantly reduced by using Q tomography. The inversion of the synthetic dataset and the bootstrap dataset behaved similarly, leading us to conclude that the biases shown from the synthetics were genuine and that the synthetic model used was appropriate. The successful reconstruction of the synthetic Q model also indicates that the reduction in variance in adopting the Q model is not simply a result of simplifying the model variance. We were able to produce error estimates for the individual parameters, along with these biases. In applying this method to the UK dataset we found: a depth-dependent, frequency-independent Q, ranging from ~920 at the top of the crust to ~5700 at the base; a multi-segment apparent geometrical spreading function; the majority of stress-drops range from of the order 0.1 to 10 MPa from $2 < M_L < 4.7$ with a clustering of similar stress drops for events of close proximity; a defined relationship between M_w and M_L, and site corrections which correlate with surface geology at a regional scale. We have found that the application of the method described above to a large database of micro-earthquake recordings can provide stable and robust estimates of the source, path and site parameters. Some assumptions still exist, such as the common source model and single frequency dependence or independence of Q for all events. The propagation of error between the steps in our inversion method is clearly a limitation. For instance, we rely on the Q model to constrain the inversion to provide robust estimates of the moment and the stress drop, but in doing this we may bias the results, if for instance the Q model is incorrect. The extensive synthetic and bootstrap testing showed that we should, however, be able to obtain the correct Q model, as long as it is not more complex than a layered medium. It is, however,

impossible to perform such a decomposition of micro-earthquake records without some assumptions. A further limitation of our model may be that we cannot consider the effects of non-linear soil response to larger earthquakes. This may mean that our predictions may be somewhat over-estimating the ground-motion for very large magnitude events with ground motions exceeding ~1ms^{-2} (e.g., Beresnev and Wen, 1996). In this case, the physical limits of the propagating medium may actually limit the maximum ground-motions, a scenario for which we cannot test using our dataset. However, in the case of the UK, a significant earthquake would be considered to be around 5-6M_w, a magnitude at which non-linear response would probably not affect the ground motion except at very short source-to-site distances for which strong shaking levels could be expected. An important caveat is that whilst our assumptions may not be precisely correct (as shown by the synthetic testing), the ability to model the Fourier spectrum of a catalogue of earthquake recordings remains good regardless, to a certain degree, of the assumptions made. We simply obtain a different model which fits the data equally, or at least with negligible difference in fit. In terms of seismic hazard analysis what is important is that the results of these studies are taken as an inter-dependent package: once we begin to independently vary several parts of the model, the covariance which holds the whole model together breaks down, and large uncertainties will develop. However, the implications of this non-uniqueness are that in scaling to higher magnitudes we may extrapolate these relations in the wrong 'direction'. More constraints are required which are valid across the magnitude range of interest if we wish to do this. Whilst the non-uniqueness is a hindrance of the weak-motion database, we can, nevertheless, take away important seismological observations and use them to carefully calibrate the attenuation relations derived directly from strong-motion recordings; or to perform stochastic simulations as long as the covariance is carefully considered.

3.9 Acknowledgments

This project is funded by the Natural Environment Research Council and the British Geological Survey. The data used in this project was made available by the British Geological Survey. The figures in this paper were made using GMT 4 (generic mapping tools) software written by Wessel and Smith (1998). Parts of the code used in creating the inversion tools were provided by Pozo (2004) and Press et al. (1997). We would like to thank Art McGarr, Stéphane Drouet and an anonymous reviewer for their helpful comments.

Chapter 4

4.1 Preface to Chapter 4

The following chapter is an edited version of a paper currently under review for publication in the Bulletin of the Seismological Society of America. It may therefore change somewhat prior to publication. It may be treated as a self-contained article, but ultimately serves as a a follow-on to the previous chapters, building on the methodology and results already presented.

The previous chapters presented the inversion method and the spectral model we derived using British data. Whilst the results of that chapter may be used to predict the spectral content of an average earthquake from Great Britain, we did not proceed to use this information for stochastic modelling. Stochastic modelling facilitates the conversion of a spectral model into what is commonly termed an 'attenuation relation', although a more logical term, commonly used by seismologists, is a 'predictive ground motion relation'. Attenuation relations predict the response of a simple harmonic oscillator, typically damped at 5% of critical. This is in order to simply but effectively provide a measure of a buildings response to ground shaking. The reason behind the conversion into these attenuation relations is therefore primarily for engineering applications: where the interest is in the structural response to earthquakes.

Before approaching stochastic modelling it is first necessary to address an issue common with geophysical problems: is the solution we have derived, in this case the spectral model in chapter 3, applicable outside the parameter ranges defined by the dataset. This is necessary due to the reliance of the stochastic model on the input spectral model. If the spectral model is wrong at high magnitudes of ground-shaking, then the stochastic model, and hence the predictions of peak ground motion, will also be wrong.

Whilst in Chapter 3 we were able to derive a robust spectral model for Great Britain, we were not able to test that the model is applicable to magnitudes beyond those used in our dataset ($2 \leq M_L \leq 4.7$). In order to test the applicability of the spectral model outside the magnitude range defined by the data, we apply the method to a seismically active region. We can then use only the small-magnitude data from this region, and using the resulting model, directly compare the resulting prediction of large-magnitude earthquakes with empirical observations. In the following chapter we therefore test the hypothesis 'is it possible to predict strong ground motion using only small-magnitude records'. We do this using an 'ideal' dataset: with excellent quality high density recording on digital broadband instruments. In addition we use borehole located instruments in order to minimise the effect of the site response, that can significantly influence our results, as shown in Chapter 2. These tests will show the feasibility of the extrapolation from weak- to strong ground-motion. An ideal region, that meets all the requirements for testing the hypothesis is Japan. The chapter is entitled 'a comparative study on attenuation and source scaling relations in the Kantō, Tokai and Chubu regions of Japan, using weak and strong ground-motion datasets'.

A comparative study on attenuation and source scaling relations in the Kantō, Tokai and Chubu regions of Japan, using weak and strong ground-motion datasets.

Benjamin Edwards[1], and Andreas Rietbrock[1].

1. Department of Earth and Ocean Sciences, University of Liverpool, UK.

Keywords: attenuation, Q, geometrical spreading, magnitude, moment, site response, kappa, ground motion, Japan, source scaling.

4.2 Abstract

Attenuation relations are derived for central Japan (spanning the Kanto, Tokai and Chubu regions) using recordings of small earthquakes ($2.0 \leq M_{JMA} \leq 4.0$), and moderate to large-magnitude earthquakes ($3.0 \leq M_{JMA} \leq 7.2$). We independently analyse both small- and large-magnitude data to provide an insight into the use of attenuation relations derived in regions of low seismic activity. A strong correlation is found between the attenuation parameters derived from each dataset. We find that Q is strongly depth-dependent, and that apparent geometrical decay increases with increasing hypocentral distance. This is modelled by using a three segment decay function, with the initial decay forced to 1/R. Moment magnitudes are close to the published M_{JMA} magnitude, but are, on average, slightly higher. An increase in stress drop with magnitude is required in order to model both the small- and large-magnitude datasets. Alternatively we show that a constant stress drop model is suitable to model the response spectra of both small- and large-magnitude earthquakes when considering the saturation of the source-corner frequency due to a static site filter such as f_{max} or kappa.

We test our ability to predict strong ground-motion by using our attenuation and source scaling relations derived from the small-magnitude recordings to stochastically simulate PGA, PGV and 5% damped response spectral ordinates over a range of magnitudes and distances. The residuals of this simulation are found to be independent of distance and magnitude. We compare our attenuation relations against other relations derived for Japan. The residuals of these relations are analysed and compared against those obtained from the model found in this study. We find that, in this study, the prediction of strong ground-motions is possible using only small-magnitude data, and that the validity of the prediction extends across all magnitudes available for comparison ($2.0 \leq M_{JMA} \leq 7.8$). Using an alternative published predictive relation for Japan PGA is significantly over-estimated for small magnitude earthquakes.

4.3 Introduction

The design of safety-critical facilities; high-rise commercial or residential structures requires the consideration of seismic hazard even in regions of low seismic activity. In evaluating the seismic hazard of an area we require ground motion equations, which, for a defined source

magnitude and location will predict the degree of ground shaking that will occur. These equations are often empirically derived (e.g., Akkar and Bommer, 2007; Ambraseys, 2005). However, the validity of predictive ground-motion equations outside of the magnitude range for which they are derived is a contentious issue. As such, in regions of low seismic activity data are taken from a non-regional (i.e., continental or part-continental) scale in order to provide validity to the equations at higher magnitudes. In doing this the possibility of regional source scaling and attenuation is neglected, which is particularly important considering the differing tectonic regimes which are used to provide a sufficient quantity of high-magnitude data.

Studies have shown a dependence of statistically derived ground motion equations on magnitude (e.g., Bommer et al., 2006; 2007). It is argued that the complexity of this magnitude dependence means that the propagation of predictive equations beyond the boundaries of the dataset used in its derivation is not possible. Alternatively the problems encountered in scaling weak- to strong-ground motion equations may be due to the non-unique solution. As a result the solution may not be valid when new data (e.g., data from larger magnitude earthquakes) is added. By including all magnitudes for which we wish to predict in the dataset we force the solution to be valid over all magnitudes. In regions of low seismicity this approach is not feasible without taking data from outside the study region.

An alternative method to formulate predictive ground motion equations is to find the physical attenuation properties and source scaling relations of the study region. Assuming that these physical properties are valid for all seismic wave-fields we can then simulate a range of earthquakes of different magnitudes using a stochastic method (e.g., Boore, 2003). In this way we have solid physical foundations upon which to apply the random vibration stochastic method. If attenuation is magnitude dependent, we should see this in the physical attenuation properties derived from datasets comprising of different magnitude earthquakes. It is therefore the aim of this study to compare the attenuation parameters, such as Q, apparent geometrical spreading and site terms derived for the same region from independent small- and large-magnitude data. We define small-magnitude data to be records of earthquakes commonly found in areas such as the UK and north-western Europe (e.g., $2.0 \leq M_{JMA} \leq 4.0$), whereas we term large-magnitude data to be from significant earthquakes relative to a low level of seismicity (e.g., $M_{JMA} \geq 4.0$). In addition to directly comparing the attenuation parameters (Q, geometrical spreading, etc.) derived from the different datasets the small-magnitude database will be used to simulate ground-motions (e.g., Boore, 2003) and the corresponding 5% damped response spectra of large earthquakes. This enables us to test the feasibility of using only small-magnitude data to predict strong-ground motion. There are numerous considerations required when postulating how the results of such a hypothesis may be transposed to other regions. Such considerations might include the quality of the data, the distribution and density of data coverage, all of which are very good in Japan, which has a high quality dense seismic

network. As far as possible we choose to limit the data we use to that commonly found in areas where the method would be applicable, such as the UK. For instance we use only the vertical component of recording, as 3 component records are not always available in old low density networks. Many more tests are required to fully backup the use of this method in other regions: for example its application to similar regions of high seismicity, where the prediction of PGA from small-magnitude data can be tested against empirical observations.

4.4 Data

We obtain large-magnitude data from Japan's NIED (National Research Institute for Earth Science and Disaster prevention) Kik-Net network (Figure 4.1, Figure 4.2).

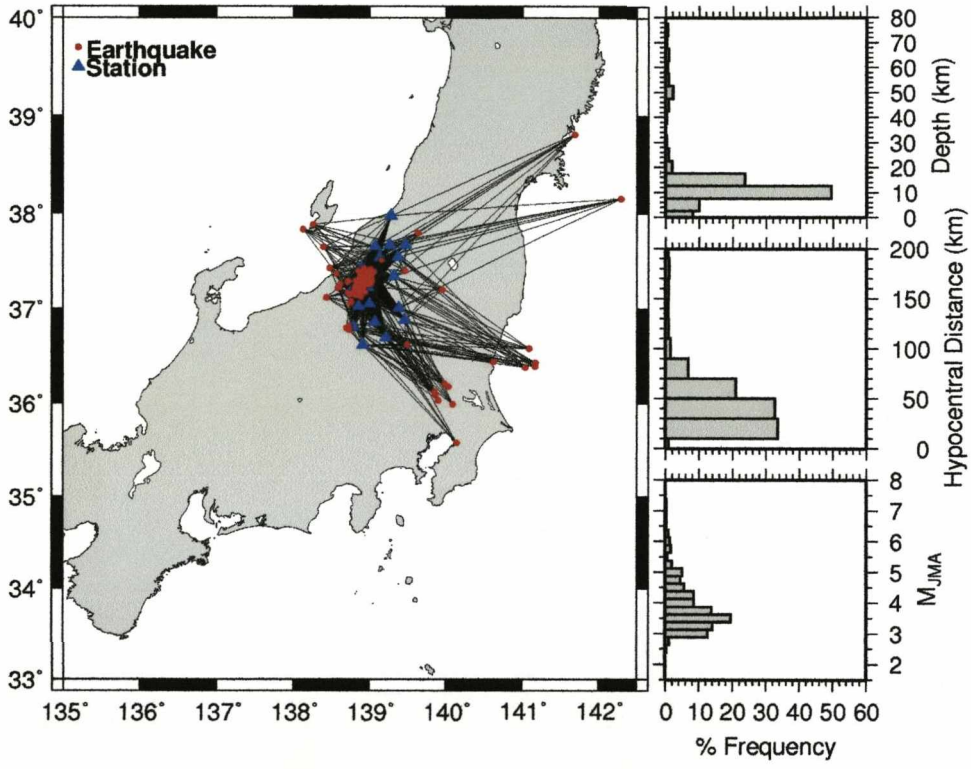

Figure 4.1: Kik-Net stations (triangles) used and events defined as large-magnitude (circles) and their respective ray paths for large-magnitude dataset 'A'. Only events, stations and ray-paths used in the inversion are shown.

This network consists of densely spaced three component dual surface/bore-hole accelerometers (see Aoi et al. (2004) for a detailed overview of the network and the installation characteristics). In this study we use only the borehole records. Whilst this is contrary to the previous comment that we will emulate the older low density seismic network, which are unlikely to have borehole installations, the borehole record provides a record that has a higher signal to noise ratio and will not be affected as much by strong site effects. Further work will use the surface record, and test the implications of the near-surface site response. The small-magnitude data were obtained from the N.I.E.D.'s Hi-Net network (Figure 4.3).

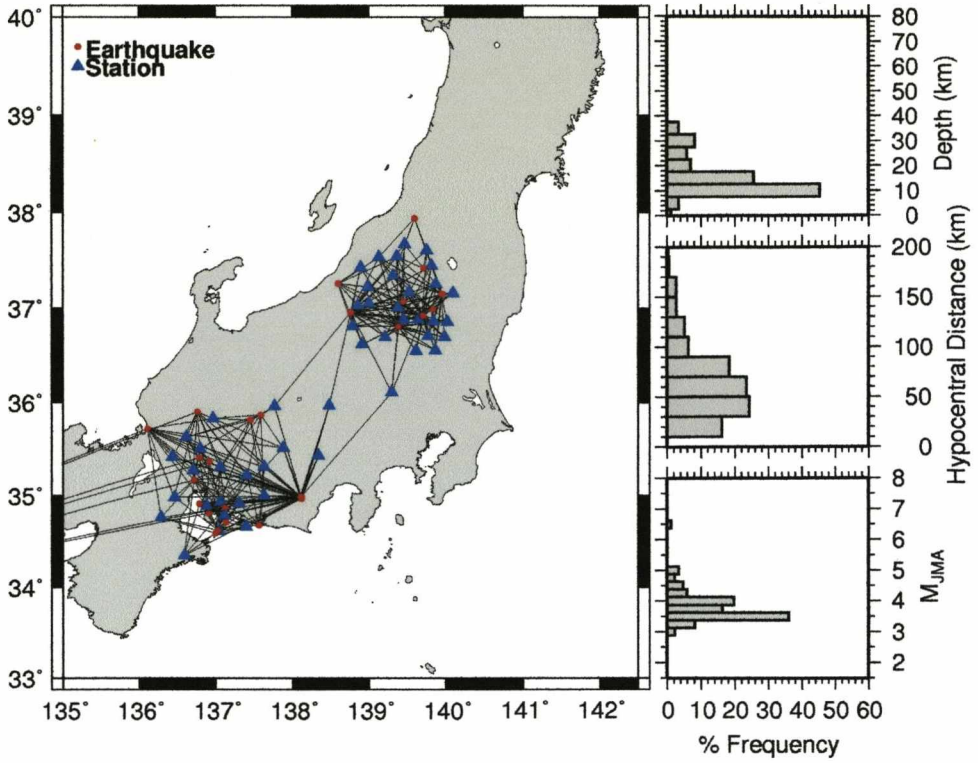

Figure 4.2: Kik-Net stations (triangles) used and events defined as large-magnitude (circles) and their respective ray paths for large-magnitude dataset 'B'. Only events, stations and ray-paths used in the inversion are shown.

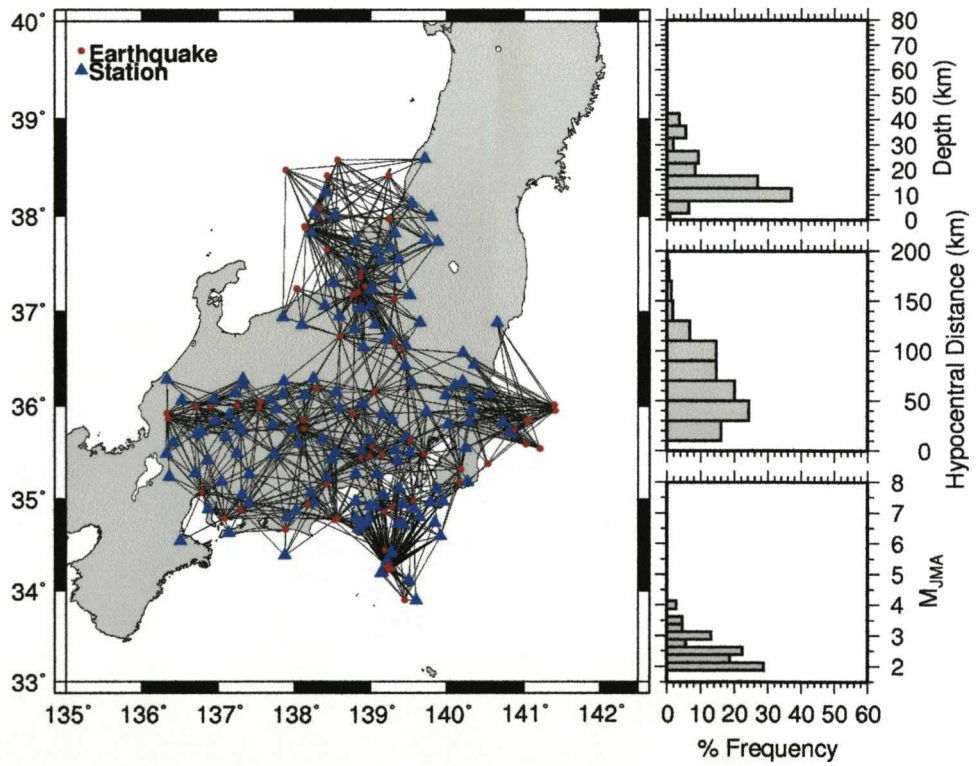

Figure 4.3: Hi-Net stations (triangles) used and events defined as small-magnitude (circles) and their respective ray paths used in the inversion. Only events, stations and ray-paths used in the inversion are shown.

This network uses high gain 1s seismometers to record low magnitude shaking. Instruments are borehole located at 100m or more in depth. Events occurring broadly within the rectangular region 32 to 39 degrees latitude and 136 to 142 degrees longitude and of hypocentral depth of no greater than 80km were selected from both networks. The large-magnitude data were split into two subsets, 'A' and 'B'; dataset A consisted of 252 events of JMA (Japan Meteorological Agency) magnitude (M_{JMA}) 3.0 to 7.2 recorded at 18 stations between 2002 and 2007; dataset B consisted of 90 events of M_{JMA} 3.0 to 6.4 recorded at 57 stations between 2000 and 2002. The small-magnitude data were limited to events of magnitude 2.0 to 4.0 recorded at 128 stations between 1st December 2006 and 5th January 2007.

We imposed several restrictions on the use of events from each dataset. An on-scale signal window of 8 seconds around the maximum amplitude was required. The signal window was further refined by selecting the 5 to 95% energy integral of this initial window. Noise estimates were computed through the windowing of the first 4 seconds of the record. Both signal and noise windows were then tapered using multi-taper algorithms (Lees and Park, 1995 and Park et al., 1987) with 5-3π prolate tapers. Finally a Fast Fourier Transform was applied to obtain the Fourier velocity spectrum of each record. The maximum bandwidth of each record was limited based on the recording instrument (Kik-Net to 0.01-25Hz and Hi-Net to 1-25Hz). We then searched for the maximum bandwidth between these values for which all spectral points are at least three times greater than the noise at that frequency. For the Hi-Net instruments we carefully remove the instrument response from the seismogram using data provided by the N.I.E.D., resulting in a record of ground velocity. For the Kik-Net network the recorded seismogram is proportional to the ground acceleration across the frequency band used (0.01-25Hz) so no instrumental correction is applied. The Kik-Net record is, however, integrated with respect to time giving a record of ground velocity. Records further than 500km from their source were discarded. The resulting data are used in the spectral and tomographic inversion and are shown in Figure 4.1, Figure 4.2 and Figure 4. 3. Large-magnitude dataset A is dominated by the Mw=6.7, 2004 Mid Niigata Prefecture earthquake and aftershock sequence (Honda et al., Internet source). As such the coverage is dominated by a smaller area than the small-magnitude dataset, which is more uniformly distributed over the study region. Figure 4.4 shows the distribution of hypocentral distance and depth against magnitude for the three datasets.

The large- and small-magnitude datasets differed slightly in terms of their hypocentral distance – magnitude distributions. The small-magnitude data, ranging from magnitude 2 to 4, was dominated by the close (up to 75km) weaker magnitude earthquakes (up to magnitude 3); while the large-magnitude datasets tended to be dominated by mid-magnitude earthquakes (3.5 – 4.5 M_{JMA}) at distances of around 20 to 50km.

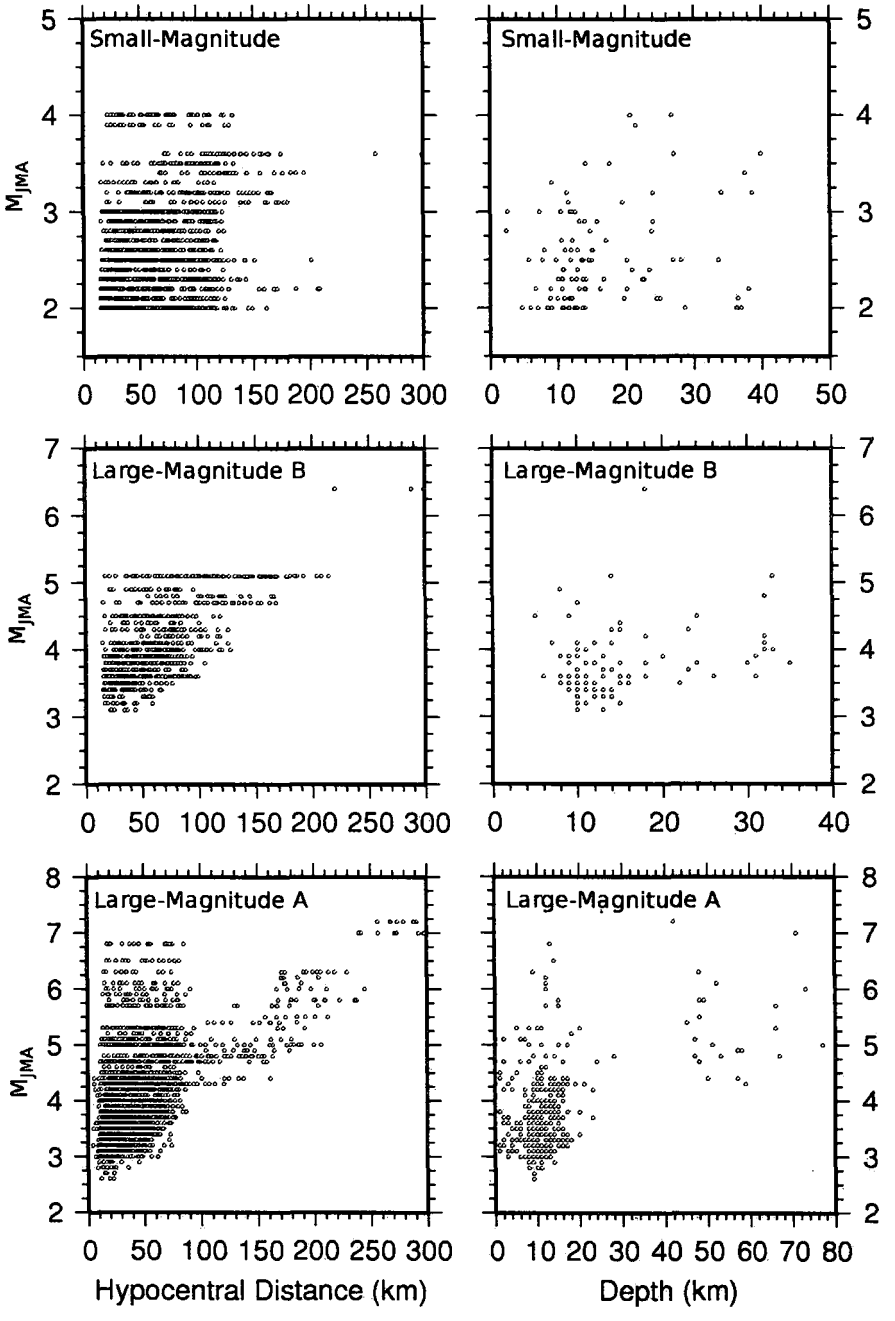

Figure 4.4: Distribution of data in terms of magnitude plotted against (left) hypocentral distance for each record and (right) source depth of each event for the three datasets used in the study.

Large-magnitude datasets A and B differ in that dataset A includes many earthquakes above magnitude 5; whereas dataset B includes only a few. Large-magnitude dataset B therefore represents earthquakes more likely to be occur in an area of low seismic activity (over a long time period). The vast majority of the small-magnitude data are recorded at distances of no greater than 150km, with the most data recorded in the first 100km. There is then a decrease in the number of records as hypocentral distance increases. This trend is independent of magnitude. In contrast, the large-magnitude data are recorded at distances which depend on

the earthquake magnitude: magnitude 3 earthquakes are recorded only a distances of up to around 50km; whereas magnitude 7 earthquakes are recorded at distances up to 300km, with a linear increase at intermediate magnitudes. The overall density of recording is higher in the first 100km from the source. These recording trends occur due to the sensitivity and trigger levels of the instruments. Most earthquakes for all datasets were located at less than 20km in depth; although earthquakes were located at depths of up to 40km for the small-magnitude dataset and large-magnitude dataset B. Earthquakes related to the subducting crust were located at up to 80km in depth in large-magnitude dataset A; but these were very sparse in quantity. Overall the depth distribution is independent of magnitude for all datasets.

4.5 Attenuation Model

We apply the method detailed in Edwards et al. (2008), which is an extension of the method of Rietbrock (2001), to the above three datasets (small-magnitude, large-magnitude A, and large-magnitude B) in order to robustly determine source, path and site terms using the following model of the Fourier velocity spectrum:

$$\Omega_{ij}(f,r)=2\pi E_i(f)\cdot B_{ij}(f)\cdot S_{ij}(r)\cdot T_j(f)\cdot I_j(f) \tag{4.1}$$

where f is the frequency, r is the hypocentral distance, $E_i(f)$ is the (Brune 1970, 1971) source model (the amplitude spectrum at the source), $B_{ij}(f)$ is the attenuation along the ray path:

$$B_{ij}(f)=e^{-\pi f t^*_{ij}}, \tag{4.2}$$

where t^*_{ij} is the frequency independent whole path attenuation operator defined by:

$$t^*_{ij}=\frac{T_{ij}}{Q_{0ij}}. \tag{4.3}$$

T_{ij} is the travel time, Q_{0ij} is the path dependent dimensionless quality factor (e.g., Rietbrock, 2001). $S_{ij}(r)$ is the amplitude decay with distance or 'apparent geometrical spreading', given by:

$$S_{ij}(r)=\begin{cases}\left(\frac{r_0}{r}\right)^{\lambda_1} & r\le r_1\\ S(r_1)\cdot\left(\frac{r_1}{r}\right)^{\lambda_2} & r_1\le r\le r_2\\ \vdots & \vdots\\ S(r_n)\cdot\left(\frac{r_n}{r}\right)^{\lambda_n} & r\ge r_n\end{cases} \tag{4.4}$$

$T_j(f)$ is the site transfer function at the recording station and $I_j(f)$ is the instrument response function. $I_j(f)$ as we removed the response of the instrument where applicable.

In deconvolving the spectral dataset we aim to extract a layered Q model, an apparent geometrical spreading function, source corner frequencies, seismic moments, and the site response at the receiver depth. In Edwards et al. (2008) we found that a strong covariance existed between $B_{ij}(f)$, the path dependent measure of attenuation and $E_i(f)$, the source spectrum (more specifically the source corner frequency), when the inversion is insufficiently constrained. In this situation we also found that a strong covariance existed between t*, a measure of path attenuation, and the signal moment (a product of the long period plateau value at the source, the geometrical spreading function, and a site amplification term) as well as between the signal moment and the frequency dependence of Q. In order to reduce the effect of the strong covariance between the attenuation parameters we use the method detailed in Edwards et al. (2008) which can be summarised in four stages: first an initial inversion of the spectra is performed for three defining parameters – the source corner frequency, the signal moment (the frequency independent far-field amplitude) and t*. From this we keep only the estimates of the t* values. Using the t* values we then tomographically reconstruct a layered Q model; this is similar to how travel time estimates may be used to reconstruct velocity structure. In a third stage we compute theoretical t* values for each of our records using ray-tracing through the the new Q model. The remaining spectral parameters (fc: source corner frequency, and the collective amplitude parameter, the signal moment) are then inverted for whilst fixing the t* value as computed. We fit the spectra in the log-space to avoid influence from the site response, a(f), for which we do not directly invert, but construct from the residuals of the inversion:

$$\log(a_j(f)) = \frac{1}{N}\sum_{i=1}^{N}\log(\theta_{ij}(f)) \quad , \tag{4.5}$$

where $\theta_{ij}(f)$ are the residuals (data/model) at discrete frequencies f over all events, i, for a given site, j. In this case, since we are using borehole records, the site response is only applicable to the depth at which the installation is located. There will be a further transfer function to account for the upper 100 or so meters of the crust that we cannot account for in this case and is beyond the scope of this paper.

In terms of the choice of the fit function, we tested other minimisation functions but they prove less stable (as shown later). We find that the log-space fit also significantly reduces the covariance between t* and the signal moment present in the initial inversion (i.e., prior to the Q structure constraint). The strong trade-off between the inversion parameters is controlled by the Q model constructed from the initial estimates of t*. It was shown in

Edwards et al. (2008) that the method could successfully decouple the attenuation parameters and that errors were available through bootstrap testing. It is important to understand that there is also a trade-off between the complexity of the Q model and its robustness. Additionally our choice of Q model will affect the other inversion parameters: over-simplification of the Q model will lead to bias in the remaining parameters to correct for that simplification, whilst too much complexity with insufficient data (quantity or quality) will lead to unstable inversion results. Our choice of a layered Q model lies between the simplest model of Q: a homogeneous half-space model; and a 3D Q model. To test that the Q model is stable we test different starting models. This of course does not give any information regarding bias of the solution. In the case of a biased solution we also bias the remaining spectral parameters as they are dependent on the t* value computed from the Q model. We test for errors and modelling bias using bootstrap testing. A limitation with the current method is the assumption that for each event the log-mean site response, a(f), is nil. This may cause problems if there is a common form to the site response, but as we are using borehole records this assumption should be better fulfilled.

Having successfully extracted the spectral shape parameters, defined by the source corner frequency, the t* value, and the signal moment we then proceed to separate the signal moment into the long period plateau value at the source, mean site amplifications and the distance decay function. We define the signal moment as:

$$\hat{\Omega}_{ij}(r) = \Psi_i \cdot A_j \cdot S_{ij}(r) \quad . \tag{4.6}$$

where $S_{ij}(r)$ is given by (Equation 4.4), A_j is a site amplification parameter independent of frequency, and ψ is the long period plateau value at the source. The reader should refer to Edwards et al. (2008) for a full description of the methodology. An important aspect of this study is that the datasets are treated as independently as possible in order to assess whether the small-magnitude parameters will provide a viable estimate of strong ground-motion. This means that although inter-dependence of data may exist implicitly, we attempt to use the results of the small-magnitude data without consideration of, or use of, the large-magnitude data to predict strong ground-motion. This allows comparison with situations where the large-magnitude data is not available. For this study we also used only the vertical records of ground motion as this is the most common orientation in older single component seismometer networks (e.g., in the UK). This will allow us to assess the validity of using the vertical records to predict PGA and PGV in circumstances where only the vertical component of recording is available. Whilst this limits the study somewhat, we feel that it is important to simulate as closely as possible the environment in which this method may be applied in regions of low seismicity. Further work will be undertaken to assess the use of the horizontal component of recording for similar applications where newer 3 component recording is available.

4.6 Inversion Results

4.6.1 Q Structure

The Q structure was inverted for using a four layer structure following the procedure in Rietbrock (2001). Nodes were located at 1.5km, 6km, 10km and 34km, with linear interpolation between the layers. The velocity model used was simplified from a model provided by the NIED and is shown in Table 4.1 . The starting models were homogeneous Q_0 for both the strong and small-magnitude datasets. The value of Q_0 was chosen based on a visual inspection of the log(1/Q) distribution from the path variable Q inversion. Table 4.2 summarises the Q tomography results.

Depth (km)	S Wave Velocity (kms^{-1})
1.5	1.70
6.0	3.40
10.0	3.50
34.0	4.80

Table 4.1: Simple velocity model used in the tomographic inversion for Q.

	Small-magnitude Q_0			Large-magnitude A Q_0			Large-magnitude B Q_0			All
Start Q:	400	740	900	400	740	900	400	740	900	
Residual Variance	1.3e-3	5.4e-4	5.5e-4	6.7e-4	3.2e-4	3.1e-4	1.2e-3	2.4e-4	2.0e-4	
Node Depth										
1.5 km	410	620	470	460	420	390	460	560	--	470 ± 80
6.0 km	430	620	510	460	440	440	470	550	--	490 ± 70
10.0 km	630	580	440	1300	1700	1800	1700	1800	--	1200 ± 600
34.0 km	1600	1600	1800	1200	1200	1200	910	900	--	1300 ±300
Residual Variance	4.6e-4	4.8e-4	4.5e-4	2.8e-4	2.9e-4	2.8e-4	1.9e-4	1.9e-4	--	
Variance Reduction (%)	64.62	11.11	18.18	58.21	9.38	9.68	84.17	20.83	--	

Table 4.2: Recovered Q structure (to two significant figures) for both small- and large-magnitude datasets. Different starting models are shown to assess the stability of the tomographic inversion. The value '--' indicates that the node was not resolved. The starting t variance (homogeneous Q) is shown and compared with the final (layered Q) t* variance for each model. The final column shows the mean and variance of all the models at each node.*

Table 4.2 shows that the four layer tomographic Q inversion was stable, and not significantly dependent on the starting model. We observe a reduction (at least 9%) in the t* residual (observed – computed) when using the layered model. This suggests that it is reasonable to allow for the four layer level of complexity in the model. A significant difference between the large-magnitude datasets and the small-magnitude dataset can be observed at the 10km node. The difference may be due the small-magnitude data sampling a different local Q structure. However, the correlation of the two large-magnitude datasets, which themselves sample different regions, suggests that this is unlikely. A resolution problem may exist for the small-magnitude dataset or the large-magnitude datasets. Additionally, the small-magnitude dataset 34 km node has higher Q which may trade-off with the low Q in the 10km node. The final column of Table 4.2 shows, to two significant figures, the average Q and its standard deviation across all four datasets. The node at 10km has the largest variance due to the differences between the small-magnitude and large-magnitude results at this depth. However, there is clearly a depth dependence of Q: a consistent reduction of around 10% in the t* residuals is achieved by using a layered Q model as opposed to the best homogeneous Q model. Taking into account the variances we can see two distinct areas: low Q (around 400) in the upper 8km and high Q (around 1300) below around 8km. This depth should not be considered an exact boundary due to the resolution of the layers. However, we may interpret the results to show two distinct regimes: those of the upper and lower crust. Table 4.3 shows the average path-Q (T/t*) for a simplified homogeneous half-space with velocity 3.5kms^{-1} using the Q model given in Table 4.2. This allows for the comparison between Q models derived using the different datasets in this study. Figure 4.5 shows t* at different hypocentral distances using the different Q models in Table 4.2 and the velocity model given in Table 4.1.

Hypocentral Distance (km)	Small-magnitude Q_{path}	Large-magnitude A Q_{path}	Large-magnitude B Q_{path}
10	310	370	350
20	360	470	440
50	500	740	700
100	740	980	970
150	930	1100	1140
200	1060	1200	1260
300	1220	1310	1400
500	1380	1400	1530

Table 4.3: Average path-Q (T/t) at different hypocentral distances for each of the datasets for a homogeneous velocity of 3.5kms^{-1}.*

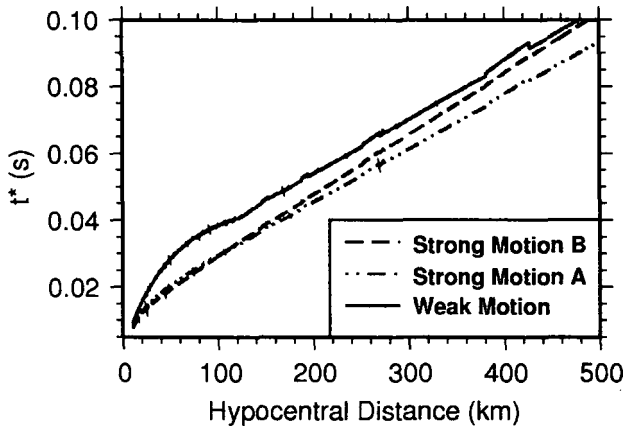

Figure 4.5: Plot of the computed t against hypocentral distance for a source at 10km depth using the Q models from Table 4.2 and the velocity model in Table 4.1. The solid line is for the small-magnitude data, the dashed line is for large-magnitude dataset B and the dash-dot line is for the large-magnitude dataset A.*

The differences between t* for the two large-magnitude Q model derived t* is negligible below 200km and increases to around 0.01 (10%) at 500km. Comparing the t* values computed for the small-magnitude derived Q model we can see that they are higher than those from both large-magnitude derived models by a maximum of around 0.01s. Unfortunately it is difficult to compare the attenuation model with other studies, as authors use different definitions of the function Q: in this case it is dependent on hypocentral distance and source depth, in other cases (e.g., Jin et al., 2000) Q is dependent on the frequency, or even shear wave velocity (e.g., Satoh, 2006). Additionally in Edwards et al. (2008) we found a strong covariance between the frequency dependence of Q and the rate of apparent geometrical decay which makes the comparison difficult of these functions problematic. However, previous published studies of Q have shown that Q≈1000 (Suzuki, 1971) in the upper crust of Honshu, central Japan, and that Q≈1000-2000 throughout the upper and lower crust (Hashizume, 1979). This is similar to the average crustal attenuation we observe, although we observe somewhat higher attenuation in the uppermost several km. Additionally these studies did not show a strong depth dependence of Q seen in these results.

4.6.2 Site Attenuation (kappa)

Figure 4.6 shows the spatial distribution of site attenuation (kappa). In the model used in this study kappa is defined as the attenuation at the site relative to the upper layer of the Q model. As a result kappa may be either positive or negative. In general we observe normal site attenuation in the fore-arc region, and mainly inverse attenuation (the amplification of high frequencies with respect to lower frequencies) as we move north-west towards the fossil back arc basin (the Sea of Japan). This could represent a lateral dimension to the Q model, or near surface attenuation due to, for instance, fluid filled fractures. Comparing the kappa values (Figure 4.6 and Figure 4.7) from the three datasets we find a correlation of similar site

attenuation at nearby stations. This is most obvious in the north-eastern part of the study area, where the three datasets significantly overlap (in terms of station recordings). The other regions do not show as prominent correlation. In general the small-magnitude dataset shows stronger attenuation or amplification in the south western region. However, there are few stations which overlap the datasets in this region: the differences here could be due to different site conditions. Additionally these stations are sampled far less by the data and the specific site terms should be used with care.

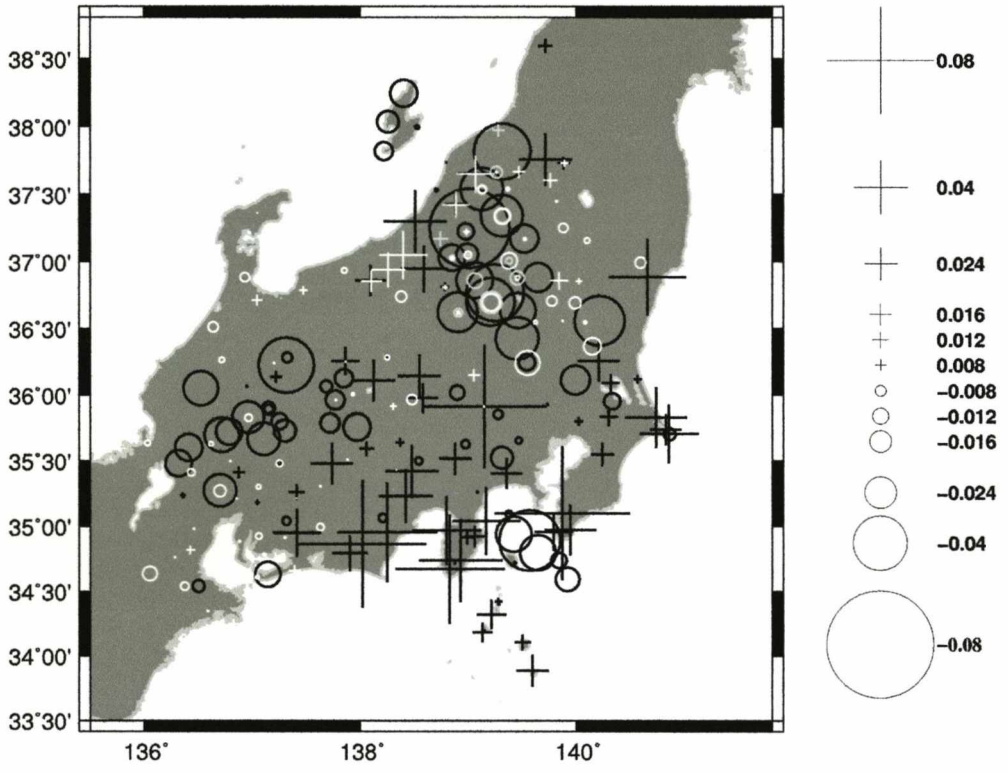

Figure 4.6: Station attenuation (kappa) values for Hi-Net stations (black) and KiK-Net stations: datasets A and B (grey and white, respectively). Crosses represent normal attenuation, circles represent inverse attenuation.

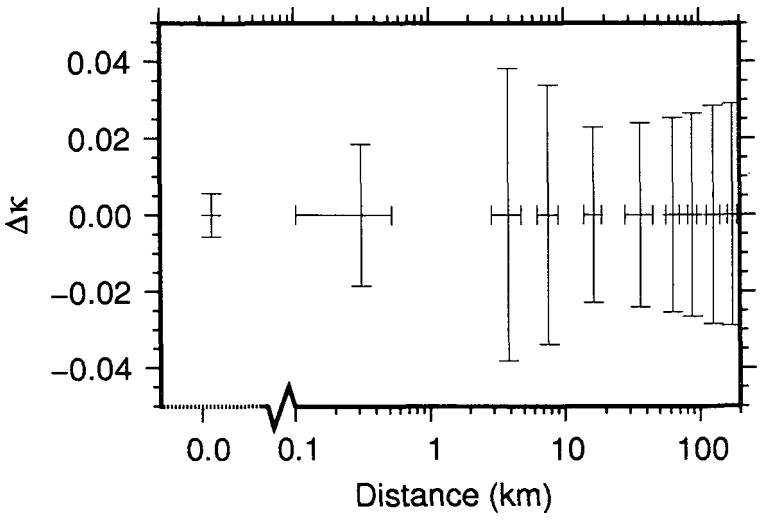

Figure 4.7: Plot of the mean and standard deviation of Δκ (the difference between the site attenuation terms at two sites) against the distance separating those two sites. The error bars show the standard deviation of the κ values over a range of separation distances, which are indicated by the error bars on the distance.

Figure 4.7 compares the site attenuation terms, κ, at all sites from the separate inversions (weak, strong A and strong B). It shows that we are able to resolve κ independently of the inversion or data used: for a separation distance of 0km (i.e., the same site) we observe a standard deviation of only 0.005. This means that more-or-less the same value is obtained regardless of the dataset we are using. For sites separated by only 100 to 500m we see a small increase in the scatter. However, this is still much lower than the scatter in the κ parameter for sites separated by over 4km. Sites within several 100m of one another are likely to share similar geotechnical and stratigraphic characteristics and so should share similar near-surface attenuation properties. The increased variability in κ with increasing separation distance corresponds to this theory and furthermore suggests that the κ, is indeed a 'site' parameter, and not an artifact of the inversion. Overall it seems that the site attenuation is well resolved and may follow a regional trend. This would facilitate the use of a simplified model for the prediction of ground motions at the surface.

4.6.3 Moment Magnitudes and Geometrical Spreading

We find a segmented geometrical spreading model (e.g., Edwards et al., 2008; Rietbrock et al., 2004) in both the small-magnitude dataset and large-magnitude dataset B. A limited epicentral distance range (mainly R < 100km) in large-magnitude dataset A make this feature difficult to resolve beyond the 110km boundary so we fix the decay value for this segment to the same as large-magnitude dataset B. Allowing λ in $1/R^{\lambda}$ to vary with no segmentation (i.e., a constant rate of decay with distance) we found that λ = 1.21 for the small-magnitude data; λ = 1.52 for large-magnitude dataset A, and λ = 1.27 for dataset B. This led to an improvement in the fractional RMS fit of signal moments of 10%; 12%, and 5.6% respectively, relative to a 1/R (λ = 1.00) decay model. However, due to covariance

between the function $1/R^{\lambda}$ and the moments, the non-conformity between the rates of decay from small- and large-magnitude data; and for formality, we chose to use $\lambda = 1.00$ for the first 30km of geometrical decay. For the segmentation distances we performed independent grid-searches at 10km spacing using the small-magnitude and large-magnitude datasets. The best segmentation distance was found at 30km and 110km for all datasets, however, for large-magnitude dataset A only the first segmentation distance was resolved. Using these distances in all three datasets we then inverted for the remaining decay exponents which are shown in Table 4.4.

Dataset	0 to 30 km	30 to 110 km	110+ km
small-magnitude	1.00 (fixed)	1.22	1.83
large-magnitude A	1.00 (fixed)	1.31	unresolved
large-magnitude B	1.00 (fixed)	1.27	1.59

Table 4.4: Amplitude decay exponents for computed from the different datasets. The values are λ where the decay is given by $1/R^{\lambda}$.

It is important not to interpret these decay exponents as simply the rate of geometrical spreading: if so then beyond 110km we observe super-spherical decay, when theory suggests we observe trapped surface wave decay (something near to $1/R^{0.5}$). Instead we are describing the amplitude decay with distance after accounting for attenuation due to our Q model. It will at least include some scattering and focusing/defocusing effects; and any other amplitude decay which occurs over a wide frequency band.

Seismic moments can be calculated using the following equation:

$$M_{0i} = \frac{\Psi_i \cdot v^3 \cdot \rho \cdot r_0 \cdot 4\pi}{\xi F \Theta_{\lambda\phi}}, \tag{4.7}$$

(Brune, 1970) where M_{0i} is the seismic moment (in SI units), ψ_i is the long-period plateau value at source i (from Equation 4.6), $\Theta_{\lambda\Phi}$ is the average radiation pattern ($\Theta_{\lambda\Phi} = 0.55$ for S waves (Aki and Richards, 1980; Boore and Boatwright, 1984)), v is the S-wave velocity at the source (we assume on average v=3.5km/s), F is the free surface amplification (F = 2.0 for normally incident SH waves and a good approximation for SV) and ρ is the average crustal density (ρ= 2800kgm^{-3}). ξ is a factor to account for the partitioning of energy onto the vertical plane (e.g., Newmark and Hall, 1982). The factor was estimated by comparing values of peak ground acceleration from the vertical and vector sum of the horizontal components of the small-magnitude dataset:

$$\log(\xi) = \log(0.52) \pm 0.26, \tag{4.8}$$

The mean value of which was comparable to that of the large-magnitude dataset:

$$\log(\xi)=\log(0.49)\pm 0.06 \quad (4.9)$$

The value of ξ was set to that of the small-magnitude data (0.52).Using the computed seismic moments, moment magnitudes are calculated using:

$$M_w=\frac{2}{3}\log M_0-6.03 \quad (4.10)$$

(Hanks and Kanamori, 1979) and are plotted against M_{JMA} in Figure 4.8.

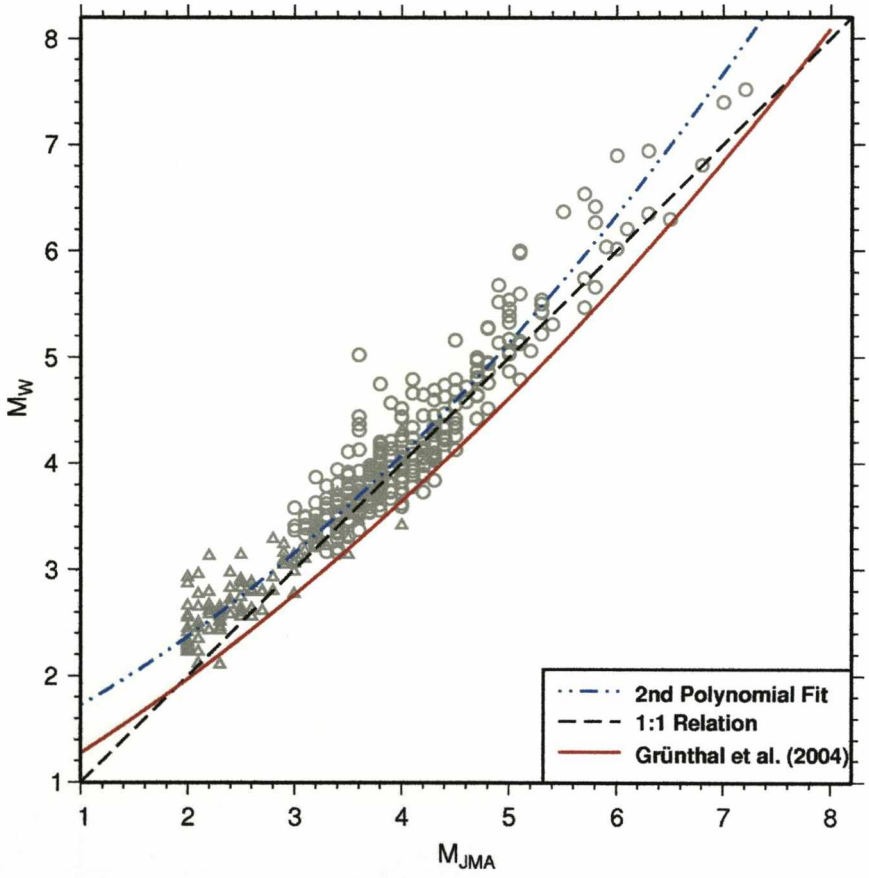

Figure 4.8: Plot of M_{JMA} against computed moment magnitude. The red dashed line shows the best fit (Equation 4.11) of Grünthal and Wahlström (2004) to a European catalogue. The dash-dot line shows the 2nd order polynomial fit (Equation 4.12) to the data. The solid line shows the linear best fit to the data. Triangles indicate events recorded by Hi-Net stations (defined as small-magnitude); circles indicate events recorded by Kik-Net stations.

The magnitudes may be separated into four groups: the small-magnitude events (M_w<3.5), dominantly those recorded by the Hi-Net stations, have a gradient of 0.8 between M_{JMA} and moment magnitudes. This sub-unity value is often encountered in studies of small-magnitude

data (e.g., Edwards et al., 2008; Drouet et al., 2006). For earthquakes with M_w less than around 4.5 but greater than 3.5, dominantly recorded on the Kik-Net stations, this low gradient continues, so it is not simply an artifact from the short period seismograph. For M_W greater than 4.5 but less than around 5.5 the gradient between M_{JMA} and M_W is around 1, such that $M_w \approx M_{JMA}$. In fact, a good approximation for $M_{JMA} > 3$ is that $M_w \approx M_{JMA}$. The increase in moment magnitude relative to M_{JMA} at low magnitudes may be due to a saturation effect as the amplitude of shaking is significantly contributed to by background noise. Grünthal and Wahlström (2004) used a second degree polynomial to model the moment magnitude – local magnitude relation for Europe. They found that :

$$Mw = 0.67(\pm 0.11) + 0.56(\pm 0.08)\, M_L + 0.046(\pm 0.013)\, M_L^{\,2}. \quad (4.11)$$

This trend is shown in Figure 4.8 and is similar in shape to the data distribution, but with a negative offset relative to the data. The L2 second degree polynomial fit (with $R^2 = 0.9$) to the data in Figure 4.8 is given by:

$$Mw = 0.439\, M_{JMA} + 0.0689\, M_{JMA}^2 + 1.22. \quad (4.12)$$

This relationship was computed assuming that the variance of M_W is approximately equal to the variance of M_{JMA} and using an orthogonal L2 fit. The shape of the function is the same as given by Equation 4.11, but without the negative offset.

4.6.4 Stress Drops

Stress drops were calculated by first estimating the radius of the rupture using:

$$r_0 = \frac{v_s k}{f_c}, \quad (4.13)$$

where v_s is the shear wave velocity at the source (we use $v_s = 3500 ms^{-1}$, the velocity at 10km depth: where most of the sources are located (Honda et al., Internet source)), and k is 0.37 for shear waves (Brune, 1970; 1971). We then take the circular rupture approximation to estimate the theoretical stress drop:

$$\Delta\sigma = \frac{7}{16} \cdot \frac{M_0}{r_0^3}, \quad (4.14)$$

(Eshelby, 1957). The stress drops are presented as a plot of moment magnitude against source corner frequency, with diagonal lines of constant stress drop in Figure 4.9. Over 95% of stress drops fall between 0.1 and 10 MPa. This range is slightly less than that found by Jin et al. (2000) of 0.01 to 100MPa. The relationship between the source corner frequency and the magnitude is important to define because the stochastic model is very sensitive to the

stress drop: changing the stress drop results in a constant, and significant, change in the predicted ground-motion over all distances. If we are to be able to predict strong ground motion using the stress drop computed from small-magnitude data we must be able to calculate the average stress drop at different magnitudes independent of the dataset used.

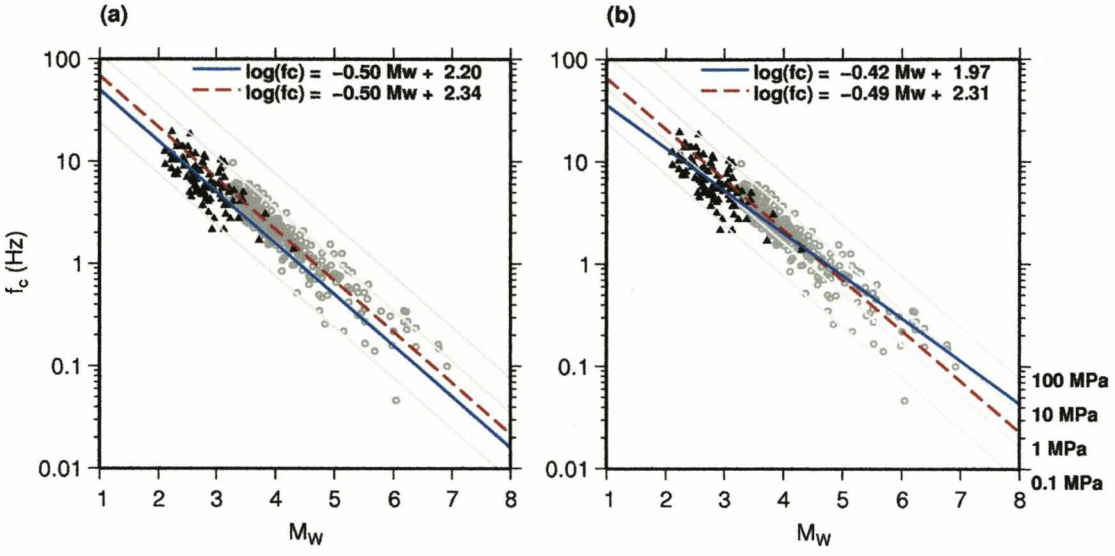

Figure 4.9: Plot of moment magnitude against source corner frequency. Triangles indicate events recorded by Hi-Net stations; circles indicate events recorded by Kik-Net stations. (a) Constant stress drop and (b) magnitude variable stress drop best fit lines are shown for the events recorded on Hi-Net (solid line) and on Kik-Net (dashed line). Light grey lines show constant stress drop, with values as labelled in (b).

The L2 trend between source corner frequency and moment magnitude using only the small-magnitude data with variable stress drop is given by:

$$\log(fc) = -0.42Mw + 1.97 \pm 0.15, (R^2=0.46). \quad (4.15)$$

In contrast using the combined large-magnitude datasets we obtain:

$$\log(fc) = -0.49Mw + 2.31 \pm 0.14, (R^2=0.81). \quad (4.16)$$

If forcing a constant stress drop we find that for the small-magnitude dataset:

$$\log(\sigma) = \log(0.91\ MPa) \pm 0.47; \quad (4.17)$$

and for the large-magnitude data:

$$\log(\sigma) = \log(2.4MPa) \pm 0.42 \quad (4.18)$$

The constant stress drop derived from the small-magnitude data is too low compared to the average stress drop of the large-magnitude datasets (around 2MPa). Figure 4.9 shows that the variable stress drop relationship derived from the small-magnitude dataset (Equation 4. 15) does look similar to that of large-magnitude datasets (Equation 4.16): an increasing stress drop with increasing magnitude. However the R^2 value for the relation derived from the small-magnitude data is relatively low and should be used cautiously.

In order to test if a magnitude dependent stress drop was necessary to model the complete range of magnitudes in this study the data was divided into three groups. The small-magnitude dataset ($2.0 \leq M_W \leq 4.2$) was retained as its own group, while the combined large-magnitude datasets were divided into two groups, one for $3.1 \leq M_W < 5$ and one for $5.0 \leq M_W < 7.5$. The average magnitudes for these datasets were 2.7, 3.9 and 5.8 respectively. Figure 4.10 shows the stress drops for the three datasets plotted in terms of the cumulative fraction of data in each group.

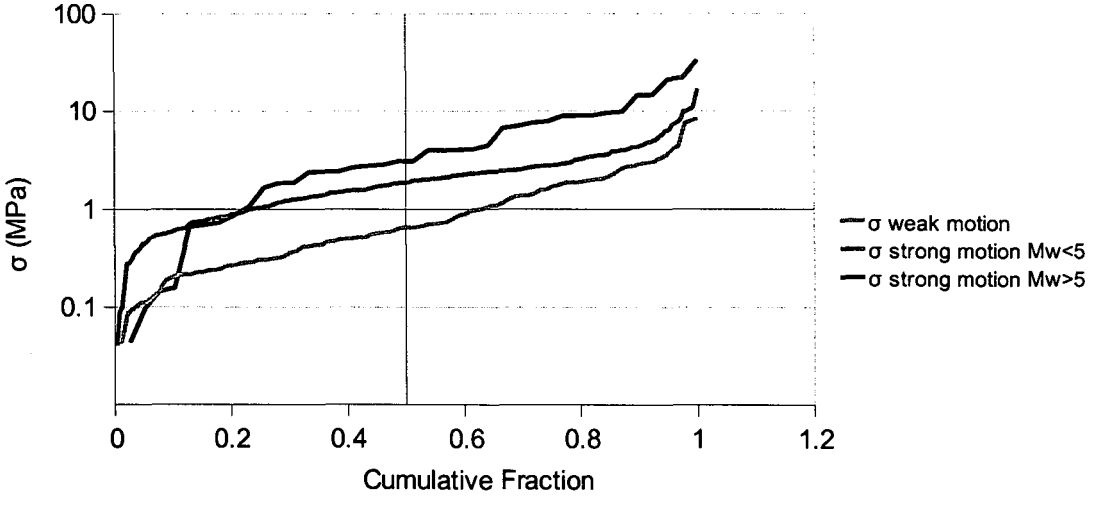

Figure 4.10: Cumulative frequency distribution of stress drop for the three data categories: small-magnitude, large-magnitude $M_W < 5$ and large-magnitude $M_W \geq 5$ described in the text.

In Figure 4.10 we observe that apart from the lower 20% of the data from each dataset, the use of an increasing average stress drop with magnitude is correct. The median stress drop for the small-magnitude dataset with average magnitude 2.7 is 0.65MPa, for that of the large-magnitude data with $M_W < 5$ (average magnitude 3.9) the median stress drop is 1.87MPa and for the group with $M_W \geq 5$ (average magnitude 5.8) the median is 5.7MPa.

Figure 4.11 compares the residual stress drop when applying the small-magnitude constant stress drop (Equation 4.17) and the small-magnitude variable stress drop (Equation 4.15) models to the stress drops from the large-magnitude data. For the data with $M_W < 5$ the

constant stress drop model tends to under-predict the large-magnitude stress drops. More specifically over 80% of stress drops for events with M_W < 5 are under-estimated using the constant stress drop model. The variable stress drop model does a far better job, with a more even distribution in the residual misfit. In this case around 60% of stress drops are under-predicted and 40% over-predicted. For the large-magnitude data with $M_W \geq 5$ the variable stress drop model does a significantly better job than the constant stress drop model. Using the constant stress drop model around 75 to 80% of stress drops are under-predicted, whereas using the variable stress drop model results in a median log residual close to zero.

This shows that the constant stress drop model derived from small-magnitude records is unsuitable for modelling large magnitude earthquakes. However, even thought he R^2 value of the variable stress drop model determined using small-magnitude data was low, it is largely successful in modelling the stress drop of moderate and high magnitude events.

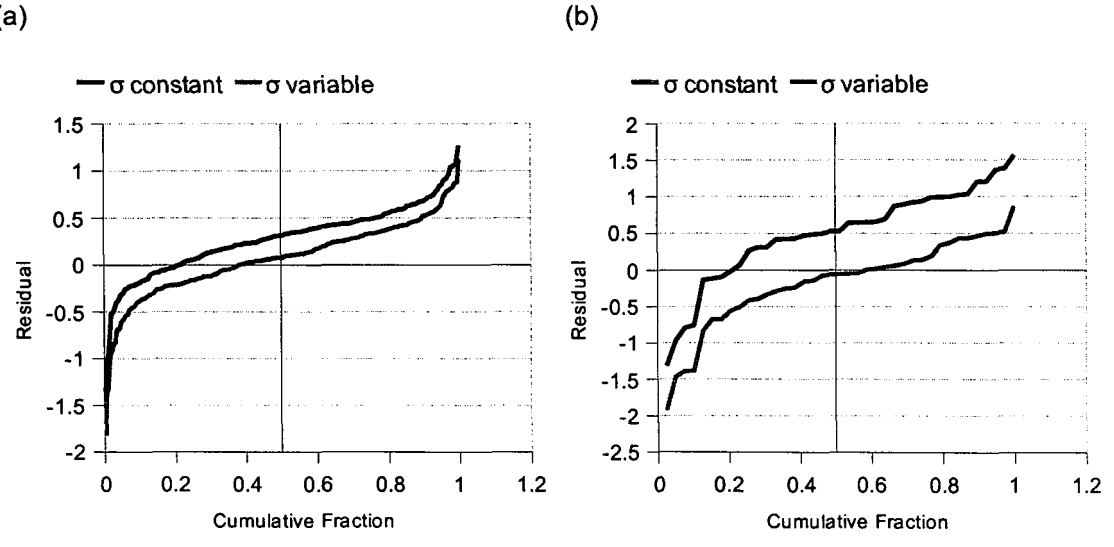

Figure 4.11: Residuals of the two small-magnitude stress drop relationships when compared to the stress drops of the combined large-magnitude datasets. The constant σ relation refers to Equation 4.17 and the variable stress drop relation refers to Equation 4.15. Large-magnitude data with (a) M_W < 5 and (b) $M_W \geq 5$.

Alternatively it can be argued that the stress drop is in fact constant: following the linear trend of the large-magnitude datasets. The lower stress drops in the small-magnitude datasets could be an artefact of the trade-off between κ (anderson and Hough, 1984) or f_{max} (Hanks, 1982) (both describing a site specific filter function) and the source corner frequency. κ and f_{max} are usually used independently and describe the high-frequency decay at around 15-25Hz often observed in empirical records. This filter is usually treated as independent of magnitude or hypocentral distance. In this study we used the κ term to describe this site filter function, so the discussion here will focus on that parameter.

However, the argument is true for either parameter.

The site specific function, defined by exp(-πfκ) can be thought of as a site specific low-pass filter with a characteristic corner-frequency. Now consider that in the case of a constant stress drop, the source corner frequency of the spectrum is highly dependent on magnitude. Therefore there exists a point at which the κ filter function removes the part of the spectrum dominated by the corner frequency: as the earthquake magnitude decreases, its source corner frequency rapidly increases. As the source corner frequency increases beyond the static corner frequency of the low-pass filter it will become indistinguishable. This will give rise to the effect observed in Figure 4.9 where the increase in the source corner with decreasing magnitude frequency slows. If we were to extrapolate back to even smaller magnitudes there would exist a saturation point at which the source corner frequency is completely mapped into the site filter's corner frequency. This is shown in Figure 4.12. The only difference between the spectra in the right hand plot is the inclusion of the filter exp(-0.04πf). In Figure 4.12 it is shown that it is very difficult to resolve the source corner frequency above around 20Hz even for the clean synthetic spectra.

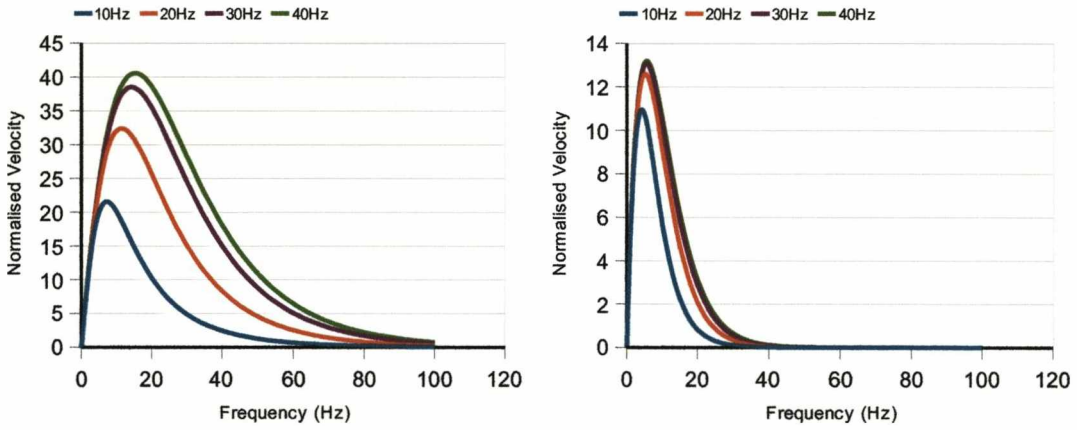

Figure 4.12: Plot of four spectra (Brune source, Q=1000, R=0km) with corner frequencies from 10 to 30Hz (see legend). Left: κ=0.0. Right: κ=0.04.

An alternative model therefore would be to use the constant stress drop from the large-magnitude dataset, but to incorporate a κ filter that is not on average zero relative to the upper layer of the crust (as assumed in this study).

4.6.5 Site Amplification

Along with obtaining the frequency dependent station correction we also invert for a mean site amplification over the entire bandwidth of the signal. If we neglect to include this amplitude correction we find a significant reduction in the fit of the long-period plateau values to our model (e.g., an increase of over 200% for the fractional misfit of the small-magnitude dataset). Figure 4.13 shows the distribution of the site amplification. We observe a trend somewhat similar to that seen for the kappa values: signal amplification relative to the

network average in the south east of the study area, becoming less as we move north-west, through to attenuation of the amplitude in the north-west. The site attenuation (κ^{-1}) values are inversely proportional to the mean amplification factor, with de-amplification in the south east, and amplification in the north west. It is not appropriate to directly compare mean site amplification values at stations overlapping the datasets, as these values are relative to each dataset's network average. Strong amplification present at some island stations in the small-magnitude dataset (possibly due to incorrectly calibrated instruments) means that other stations in this dataset require strong de-amplification to satisfy an inversion constraint. We find that there is some correlation (correlation coefficient = 0.66) between the site amplification and the site attenuation term, kappa (Figure 4.14). This correlation was also found in a UK dataset (Edwards et al., 2008). In Edwards et al., 2008, we postulated that the correlation is a reflection of the correlation of soil depth and bedrock hardness: that thick soil and soft, porous bedrock give rise to strong attenuation and strong mean amplification (resonance) at the site; whereas thin or no soil and hard bedrock give rise to little attenuation and mean de-amplification at the site. A detailed study of site conditions would be required to quantify this theory but is out of scope for this paper.

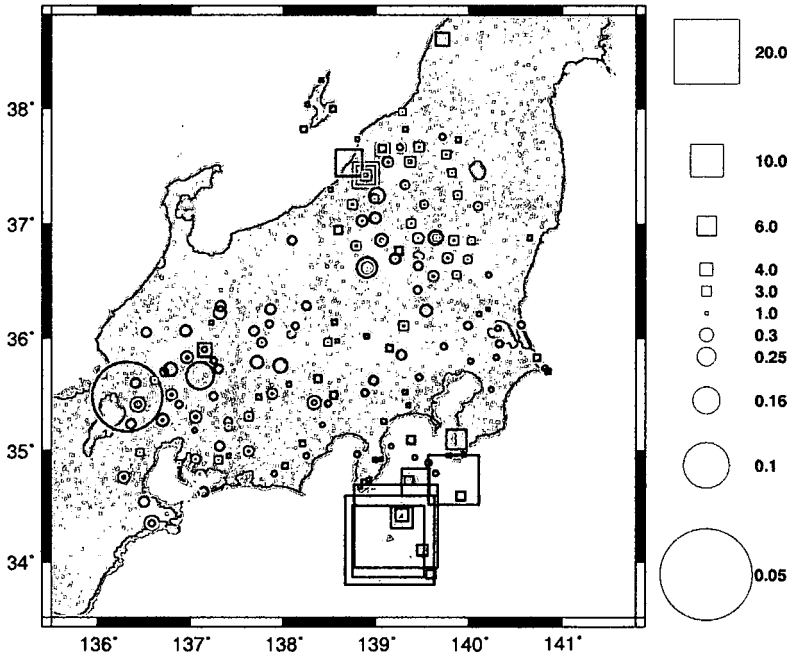

Figure 4.13: Station amplitude correction values for Hi-Net stations (black shapes); Kik-Net stations from dataset A (grey shapes) and Kik-Net stations from dataset B (white shapes). Circles represent attenuation, and squares amplification relative to the network average.

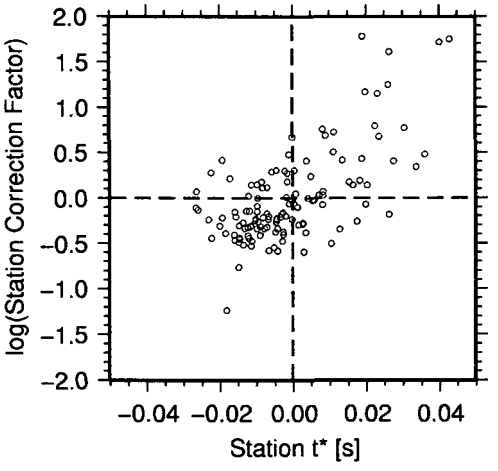

Figure 4.14: Plot of station t* (kappa) values plotted against the logarithm of the station correction factor.

4.7 Bootstrap Analysis

A bootstrap analysis allows us to derive parameter statistics by utilising the empirical distribution of data. Bootstrap analyses of (a) the real small-magnitude data, and (b) the large-magnitude dataset B were performed using 80% of all events in each dataset, with the remaining 20% being repetitions of events. Within each event we also apply the same routine; 80% of records are kept, whilst 20% are repetitions. We invert 140 of these randomised sub-datasets which equates to inverting of the order 10^5 parameters. This is repeated using a variety of optimisation functions as in Edwards et al. (2008). By comparing the results of these inversions to the results of an L2 inversion of all data in the log-space we can test the stability of the results, and any relative bias introduced due to the optimisation function chosen. Tables 4.5 and 4.6 show the 68.3% confidence interval for each parameter in terms of a percentage deviation from the final log-space L2 fit for the small- and large-magnitude datasets respectively.

Q tomography	Parameter		Stability			
		% Deviation from final results	-σ	Median	+σ	(+σ)- (-σ) *
No	t^*	-σ	-70.7	-64.3	-60.1	10.6
		Median	-7.9	-3.4	-0.5	7.4
		+σ	45.0	47.9	51.3	6.3
		(+σ)- (-σ) **	115.7	112.2	111.4	
Yes	t^*	-σ	-47.5	-36.8	-27.8	19.7
		Median	-22.7	-16.1	-9.4	13.3
		+σ	-1.1	4.3	11.7	12.8
		(+σ)- (-σ) **	46.4	41.1	39.5	
No	f_c	-σ	-36.7	-33.2	-29.7	7
		Median	-3.6	0.6	8.5	12.1
		+σ	27.6	40.1	43.9	16.3
		(+σ)- (-σ) **	64.3	73.3	73.6	
Yes	f_c	-σ	-32.3	-25.4	-21.6	10.7
		Median	-17.7	-13.6	-4.8	12.9
		+σ	0.0	10.2	21.6	21.6
		(+σ)- (-σ) **	32.3	35.6	43.2	
No	$\hat{\Omega}$	-σ	-44.0	-40.7	-36.7	7.3
		Median	-3.7	1.4	6.9	10.6
		+σ	58.1	68.5	77.2	19.1
		(+σ)- (-σ) **	102.1	109.2	113.9	
Yes	$\hat{\Omega}$	-σ	-30.3	-20.5	-13.7	16.6
		Median	1.4	9.8	16.9	15.5
		+σ	35.5	49.6	63.7	28.2
		(+σ)- (-σ) **	65.8	70.1	77.4	

*Table 4.5: Bootstrap analysis results for the log space L2 fit of the small-magnitude data. Reading up-down the rows are the parameter variances (in terms of % deviation from the full Log L2 fit results). Reading columns left-right are the stability of these variance estimates. * a measure of the reliability of the variance of the parameter. ** a measure of the reliability of the inversion.*

Q tomography	Parameter		Stability			
		% Deviation from final results	-σ	Median	+σ	(+σ)- (-σ) *
No	t^*	-σ	-80.9	-75.1	-67.9	13
		Median	-2.8	1.2	6.0	8.8
		+σ	70.5	78.1	88.7	18.2
		(+σ)- (-σ) **	151.4	153.2	156.6	
Yes	t^*	-σ	-14.0	-8.3	-2.6	11.4
		Median	8.1	12.8	16.9	8.8
		+σ	30.2	37.1	48.0	17.8
		(+σ)- (-σ) **	44.2	45.4	50.6	
No	f_c	-σ	-32.7	-31.3	-25.2	7.5
		Median	-4.7	-0.6	4.6	9.3
		+σ	57.4	67.3	98.4	41
		(+σ)- (-σ) **	90.1	98.6	123.6	
Yes	f_c	-σ	-10.1	-0.9	4.0	14.1
		Median	9.2	14.7	20.4	11.2
		+σ	26.4	39.4	53.7	27.3
		(+σ)- (-σ) **	36.5	40.3	49.7	
No	$\hat{\Omega}$	-σ	-36.8	-33.6	-29.9	6.9
		Median	2.7	7.1	12.4	9.7
		+σ	63.1	72.7	87.3	24.2
		(+σ)- (-σ) **	99.9	106.3	117.2	
Yes	$\hat{\Omega}$	-σ	-37.5	-31.5	-26.6	10.9
		Median	-10.4	-6.7	-3.1	7.3
		+σ	17.7	22.5	29.1	11.4
		(+σ)- (-σ) **	55.2	54	55.7	

*Table 4.6: Bootstrap analysis results for log space L2 fit of the surface recorded large-magnitude dataset B. Reading up-down the rows are the parameter variances (in terms of % deviation from the full Log L2 fit results). Reading columns left-right are the stability of these variance estimates. * a measure of the reliability of the variance of the parameter. ** a measure of the reliability of the inversion.*

These statistics are themselves given confidence intervals based on their variance over the 140 datasets: we find that the statistics are stable relative to the individual parameter error. The values in bold give the range between the 68.3% confidence intervals of the parameter error. They therefore represent a quantitative measure of the inversions ability to stably resolve the respective parameter: a lower range represents a more robust solution. These values are summarised in Table 4.7 for the small-magnitude dataset and Table 4.8 for the large-magnitude dataset.

Fit	$\Delta\sigma(\hat{\Omega})$	$\Delta\sigma(t^*)$	$\Delta\sigma(f_c)$
Log L2	70%	41%	36%
Log L1	64%	38%	41%
L1	115%	157%	109%
L2	219%	210%	110%
Maximum-Likelihood	342%	102%	63%

Table 4.7: Measure of inversion robustness by optimisation type and parameter for the small-magnitude dataset. Values represent the range between the median-upper and median-lower 68.3% confidence values ($\Delta\sigma$).

Fit	$\Delta\sigma(\hat{\Omega})$	$\Delta\sigma(t^*)$	$\Delta\sigma(f_c)$
Log L2	54%	45%	40%
Log L1	64%	48%	55%
L1	123%	140%	65%
L2	212%	270%	120%
Maximum-Likelihood	123%	91%	68%

Table 4.8: Measure of inversion robustness by optimisation type and parameter for large-magnitude dataset B. Values represent the range between the median-upper and median-lower 68.3% confidence values ($\Delta\sigma$).

In estimating all of the parameters the log-space L1 and L2 fits are most stable. Along with highlighting the inversion robustness and hence aiding choice of optimisation function the distributions in Tables 4.5 and 4.6 can be used as error bounds for the parameters obtained. As a result of both the synthetic and bootstrap analyses it was decided to use the log-space L2 fit when quoting the results in this paper.

4.8 The Prediction of Response Spectral Ordinates

We have so far shown that there are similarities between the attenuation models derived, suggesting that they are independent of magnitude. There are some differences but the strong covariance matrices between the attenuation parameters computed for these datasets, and for other small-magnitude datasets (e.g., Edwards et al., 2008) suggest that attenuation relations must be taken as a complete package, or carefully adjusted whilst

considering the covariance matrix. In order to test the attenuation model derived using the small-magnitude data we assumed that this was the correct model, and used it to predict ground motion from significant earthquakes using the stochastic method of Boore, 2003. In order to verify this Figure 4.15 shows the residuals of the forward modelling of the Fourier velocity spectrum for all three datasets using the attenuation parameters derived from the small-magnitude dataset and the moment magnitude relation defined by Equation 4.12.

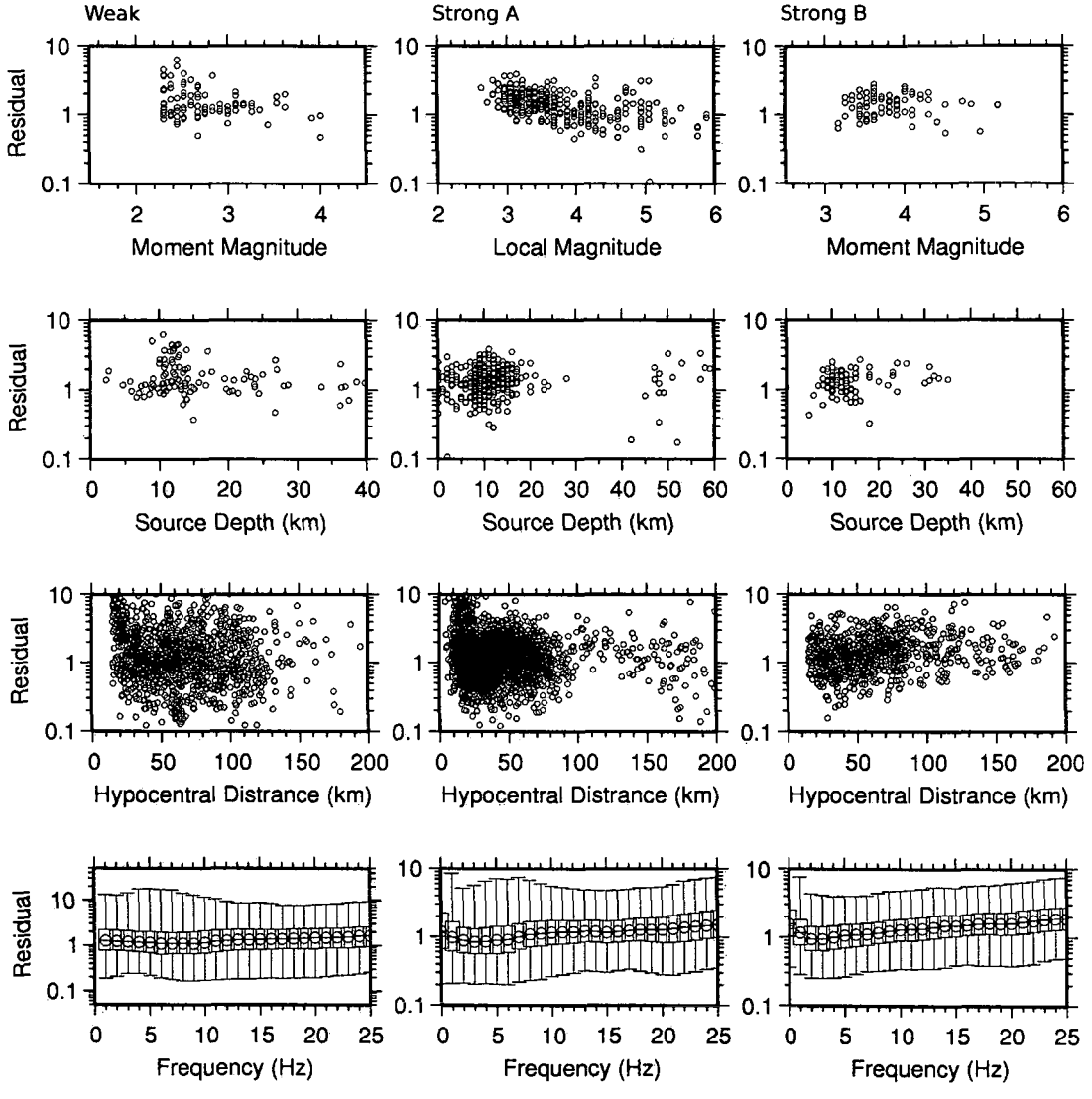

Figure 4.15: Fractional residuals (data/model) from the forward modelling of the small-magnitude attenuation parameters compared to the small-magnitude dataset, large-magnitude dataset B and large-magnitude dataset A. The box and whisker used in the plot of residuals against frequency show the 68.3 and 95% confidence intervals.

In Figure 4.15 we can see that the small-magnitude derived model does a reasonable job of modelling the Fourier spectra of the large-magnitude datasets. There are some trends in the residuals. One dependency is that of the residual against moment-magnitude for large-magnitude dataset A. In this case the amplitude of magnitude 3 to 4 events is slightly under-predicted. The majority of events of magnitude greater than 4 are correctly modelled

however. The very high frequencies (f>15Hz) tend to be slightly under-estimated for the large-magnitude records.

Using the vector sum of the horizontal components of acceleration we computed the 5% damped response spectra in ms^{-2} along with PGA and PGV values for both large-magnitude datasets used in the initial methodology. In addition PVA (peak vertical acceleration) and PVV (peak vertical velocity) along with vertical response spectra were computed from the vertical component of acceleration. Stochastic simulations were then computed using the programme SMSIM (Boore, online) using with the small-magnitude attenuation model. In the first instance this consisted of a Brune stress parameter dependent on magnitude, as this was obtained directly from the small-magnitude data:

$$\Delta\sigma = \Delta\sigma_{ref}\, 10^{0.24(M_w - M_{ref})}, \tag{4.19}$$

Where $\Delta\sigma_{ref}$ = 0.91MPa and M_{ref} = 3. This was obtained by rearranging equations 4.13, 4.14, and 4.15. However, we also later show an alternative model including a constant stress drop derived from the large-magnitude data along with a κ filter. Q is dependent on hypocentral distance (from the small-magnitude Q model in Table 4.3); and a three segment amplitude decay model (from the small-magnitude decay model in Table 4.4). The duration of shaking, T_d, was computed using:

$$T_d = T_s + T_p \tag{4.20}$$

where T_s is the source rupture duration, and T_p is the additional shaking due to the path length (hypocentral distance). For distances less than or equal to 50km the duration of shaking was only dependant on the source duration (Figure 4.16):

$$T_s = \frac{1}{f_c} \tag{4.21}$$

where f_c is the source corner frequency; estimated from the stress drop (Equation 4.19) and moment magnitude using Equations 4.10, 4.13 and 4.14. The path duration was estimated by visually inspecting a normalised density plot of the ground motion data against time and hypocentral distance (Figure 4.16). We see that for path lengths of less than 50km there is no increase in duration, but for distances of greater than 50km we find that:

$$T_p = 0.02(R - 50\text{km}) \tag{4.22}$$

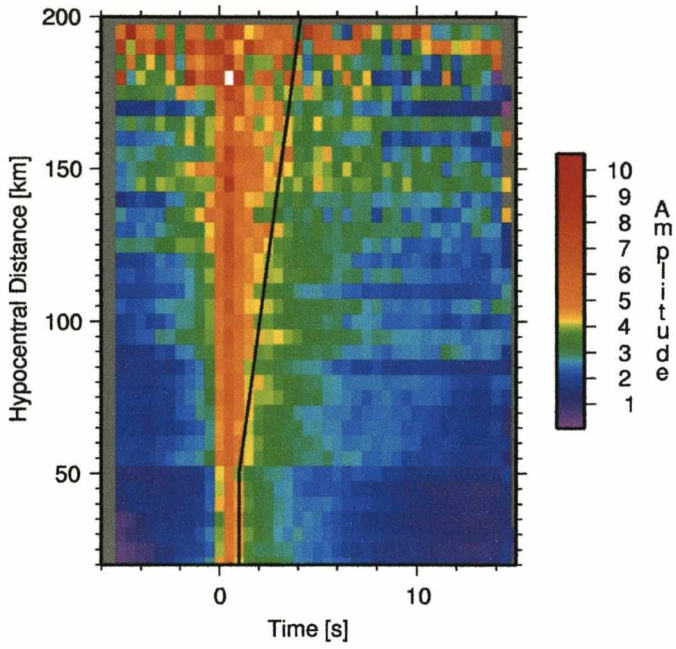

Figure 4.16: Plot of stacked RMS velocity time series from the small-magnitude database. The time axis is normalised such that the peak amplitude is always at 0 seconds. The amplitude is normalised at each distance so no geometrical decay is apparent. The black line indicates the path duration given in Equation 4.22.

As the spectral inversion used to derive this model assumed that, on average, the site response was nil, no site terms are included in the stochastic model. We apply a correction factor of 0.52 to convert from the horizontal to vertical plane (e.g., Newmark and Hall, 1982) as found by analysing the small-magnitude data (Equation 4.8) . This value is self-consistent with the value used in the moment magnitude calculation. However, the conversion between vertical and horizontal component may be dependent on magnitude and/or hypocentral distance. As the original data used to derive the attenuation model was vertically orientated any problems with this 'rule of thumb' will be mapped into the horizontal predictions. The vertical prediction will be unaffected by the choice of the factor as long as we are consistent between its use in the magnitude calculation and in the conversion of peak amplitudes. Our model yielded theoretical predictions of the PGA, PGV, and 5% damped response at 0.5Hz, 1.0Hz, 5Hz and 10Hz which are summarised in Tables 4.9 to 4.14.

	log PGA (m/s/s)				
Log Hypocentral Dist. (km)	**3.0Mw**	**4.0Mw**	**5.0Mw**	**6.0Mw**	**7.0Mw**
1.00	0.45	1.02	1.54	2.04	2.53
1.10	0.28	0.86	1.39	1.89	2.38
1.20	0.11	0.70	1.24	1.75	2.24
1.30	-0.06	0.55	1.09	1.60	2.10
1.40	-0.23	0.39	0.94	1.46	1.95
1.50	-0.41	0.23	0.79	1.31	1.81
1.60	-0.60	0.05	0.62	1.15	1.65
1.70	-0.79	-0.12	0.46	0.99	1.49
1.80	-1.07	-0.35	0.27	0.82	1.33
1.90	-1.32	-0.56	0.08	0.65	1.17
2.00	-1.56	-0.76	-0.10	0.49	1.02
2.10	-1.84	-1.02	-0.33	0.28	0.82
2.20	-2.13	-1.29	-0.58	0.04	0.60
2.30	-2.42	-1.56	-0.83	-0.19	0.38
2.40	-2.74	-1.85	-1.10	-0.45	0.14
2.50	-3.07	-2.15	-1.37	-0.70	-0.10

Table 4.9: PGA values for different magnitudes at distances spaced logarithmically between 10 and 316.2km.

	log PGV (m/s)				
Log Hypocentral Dist. (km)	**Mw = 3.0**	**Mw = 4.0**	**Mw = 5.0**	**Mw = 6.0**	**Mw = 7.0**
1.00	-1.49	-0.71	0.03	0.74	1.44
1.10	-1.63	-0.83	-0.08	0.63	1.33
1.20	-1.76	-0.95	-0.20	0.52	1.23
1.30	-1.89	-1.07	-0.31	0.42	1.12
1.40	-2.03	-1.19	-0.43	0.31	1.01
1.50	-2.18	-1.32	-0.55	0.19	0.90
1.60	-2.34	-1.47	-0.68	0.06	0.77
1.70	-2.49	-1.61	-0.82	-0.08	0.64
1.80	-2.73	-1.80	-0.98	-0.22	0.51
1.90	-2.95	-1.99	-1.14	-0.36	0.37
2.00	-3.16	-2.17	-1.29	-0.50	0.24
2.10	-3.41	-2.39	-1.49	-0.69	0.07
2.20	-3.68	-2.64	-1.72	-0.89	-0.13
2.30	-3.94	-2.88	-1.94	-1.10	-0.33
2.40	-4.22	-3.13	-2.18	-1.32	-0.53

Table 4.10: PGV values for different magnitudes at distances spaced logarithmically between 10 and 316.2km.

	Log 0.5 Hz Response (5% damped) (m/s/s)				
Log Hypocentral Dist. (km)	**Mw = 3.0**	**Mw = 4.0**	**Mw = 5.0**	**Mw = 6.0**	**Mw = 7.0**
1.00	-1.99	-0.84	0.28	1.30	2.04
1.10	-2.10	-0.95	0.18	1.20	1.94
1.20	-2.22	-1.05	0.08	1.10	1.84
1.30	-2.33	-1.16	-0.02	1.00	1.74
1.40	-2.45	-1.26	-0.12	0.90	1.63
1.50	-2.58	-1.37	-0.22	0.79	1.53
1.60	-2.72	-1.50	-0.35	0.67	1.40
1.70	-2.86	-1.63	-0.47	0.54	1.27
1.80	-3.01	-1.80	-0.60	0.41	1.15
1.90	-3.23	-1.98	-0.73	0.29	1.02
2.00	-3.44	-2.13	-0.86	0.16	0.89
2.10	-3.67	-2.31	-1.01	0.00	0.72
2.20	-3.88	-2.50	-1.20	-0.19	0.53
2.30	-4.08	-2.68	-1.38	-0.38	0.34
2.40	-4.28	-2.88	-1.57	-0.57	0.15
2.50	-4.49	-3.07	-1.77	-0.77	-0.05

Table 4.11: 5% damped 0.5Hz response values for different magnitudes at distances spaced logarithmically between 10 and 316.2km.

	Log 1.0 Hz Response (5% damped) (m/s/s)				
Log Hypocentral Dist. (km)	**Mw = 3.0**	**Mw = 4.0**	**Mw = 5.0**	**Mw = 6.0**	**Mw = 7.0**
1.00	-1.37	-0.21	0.87	1.67	2.29
1.10	-1.48	-0.31	0.77	1.57	2.19
1.20	-1.59	-0.41	0.67	1.47	2.09
1.30	-1.70	-0.52	0.56	1.37	1.99
1.40	-1.82	-0.62	0.46	1.26	1.88
1.50	-1.94	-0.73	0.35	1.15	1.77
1.60	-2.07	-0.85	0.22	1.02	1.64
1.70	-2.21	-0.98	0.10	0.90	1.52
1.80	-2.42	-1.14	-0.03	0.76	1.38
1.90	-2.61	-1.28	-0.16	0.63	1.25
2.00	-2.75	-1.40	-0.29	0.50	1.12
2.10	-2.92	-1.56	-0.45	0.33	0.95
2.20	-3.12	-1.75	-0.65	0.13	0.75
2.30	-3.31	-1.95	-0.85	-0.07	0.56
2.40	-3.52	-2.16	-1.05	-0.28	0.35
2.50	-3.74	-2.37	-1.27	-0.49	0.15

Table 4.12: 5% damped 1.0Hz response values for different magnitudes at distances spaced logarithmically between 10 and 316.2km.

	Log 5.0 Hz Response (5% damped) (m/s/s)				
Log Hypocentral Dist. (km)	**Mw = 3.0**	**Mw = 4.0**	**Mw = 5.0**	**Mw = 6.0**	**Mw = 7.0**
1.00	0.09	0.98	1.65	2.21	2.73
1.10	-0.02	0.87	1.54	2.10	2.62
1.20	-0.14	0.76	1.42	1.98	2.50
1.30	-0.25	0.64	1.31	1.87	2.39
1.40	-0.37	0.52	1.18	1.75	2.27
1.50	-0.50	0.39	1.06	1.62	2.14
1.60	-0.65	0.25	0.91	1.48	2.00
1.70	-0.79	0.10	0.77	1.33	1.85
1.80	-0.94	-0.06	0.60	1.17	1.70
1.90	-1.10	-0.23	0.44	1.02	1.55
2.00	-1.27	-0.39	0.28	0.87	1.40
2.10	-1.50	-0.62	0.06	0.66	1.20
2.20	-1.75	-0.86	-0.18	0.42	0.97
2.30	-1.99	-1.11	-0.42	0.19	0.76
2.40	-2.28	-1.38	-0.69	-0.07	0.50
2.50	-2.57	-1.67	-0.97	-0.34	0.25

Table 4.13: 5% damped 5.0Hz response values for different magnitudes at distances spaced logarithmically between 10 and 316.2km.

	Log 10.0 Hz Response (5% damped) (m/s/s)				
Log Hypocentral Dist. (km)	**Mw = 3.0**	**Mw = 4.0**	**Mw = 5.0**	**Mw = 6.0**	**Mw = 7.0**
1.00	0.53	1.24	1.82	2.35	2.85
1.10	0.40	1.11	1.70	2.23	2.73
1.20	0.27	0.99	1.57	2.10	2.60
1.30	0.14	0.85	1.44	1.97	2.47
1.40	0.00	0.71	1.30	1.83	2.33
1.50	-0.15	0.57	1.15	1.68	2.19
1.60	-0.32	0.40	0.99	1.52	2.02
1.70	-0.49	0.23	0.83	1.36	1.86
1.80	-0.68	0.03	0.63	1.18	1.69
1.90	-0.89	-0.17	0.44	1.00	1.52
2.00	-1.09	-0.36	0.26	0.83	1.36
2.10	-1.36	-0.63	0.01	0.59	1.13
2.20	-1.65	-0.91	-0.26	0.34	0.88
2.30	-1.94	-1.19	-0.53	0.08	0.64
2.40	-2.28	-1.51	-0.84	-0.21	0.36
2.50	-2.63	-1.85	-1.16	-0.51	0.07

Table 4.14: 5% damped 10.0Hz response values for different magnitudes at distances spaced logarithmically between 10 and 316.2km.

We compared the theoretical predictions with the vector sum of our observed horizontal data (e.g., Figure 4.17), assuming the moment magnitude relation defined by Equation 4.12. The fit of our theoretical model to the data is adequate: always lying within the range defined by 68.3% of the data. Some difference in the rate of decay is observed but it should be noted that this is only one particular example of an earthquake. It is more insightful to draw conclusions from the residuals of the model compared to many earthquakes.

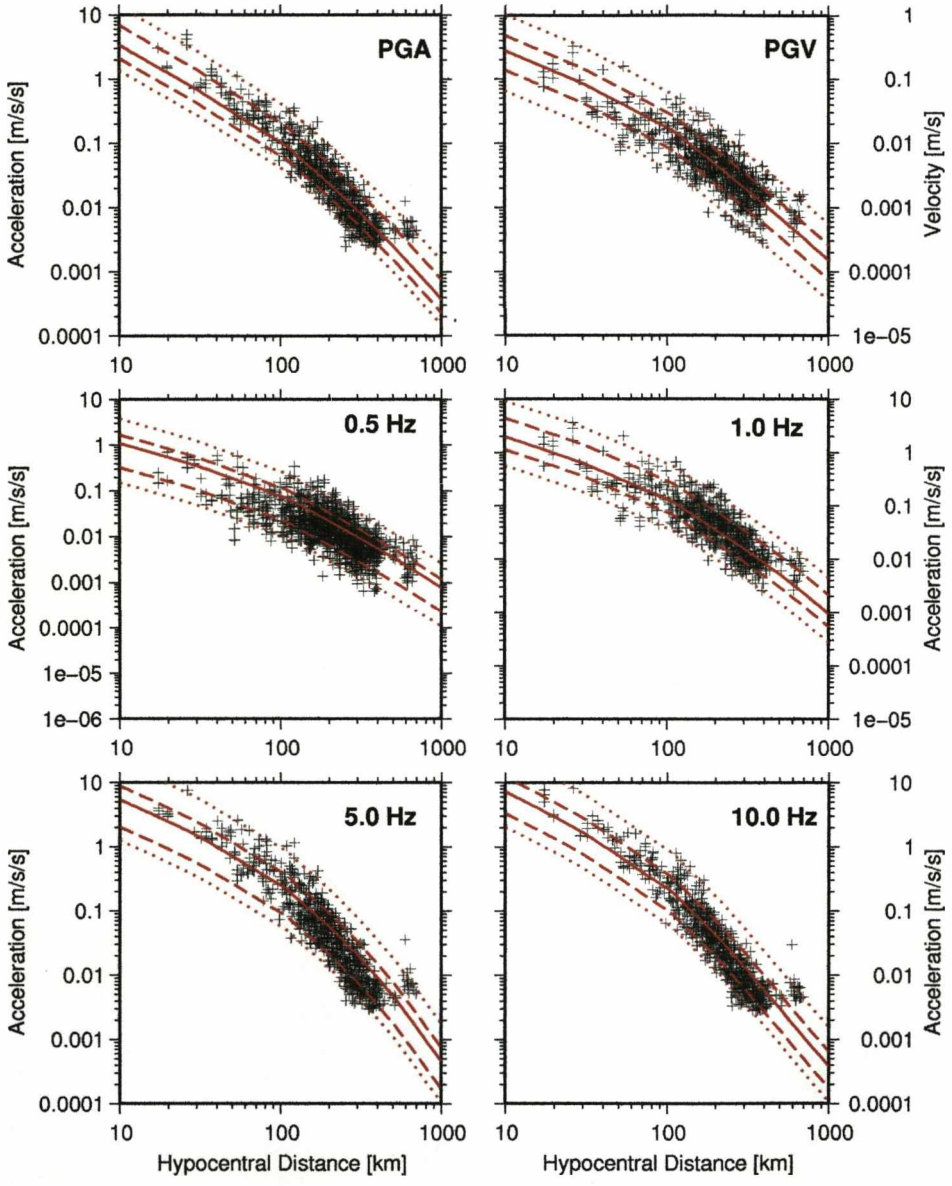

Figure 4.17: Plot of horizontal response spectral ordinates from events with moment magnitude 7.0 ±0.2. The solid line is the small-magnitude derived theoretical model; the dashed and dotted lines indicate the limits required to include 68.3% and 95% of the data (in this example only) respectively.

Confidence limits (Tables 4.15 and 4.16) were estimated after normalisation of the data with the theoretical model. This provides an empirical estimate of the error of our prediction, although it is dependent on having a complete normal distribution of data available.

We added further large earthquakes (M_w>6.5) from the study region which were previously unused (due to occurring outside our time window). Response spectral ordinates were computed for these earthquakes (as above) and they are compared to the predicted values. The residuals of all events (weak, strong A, strong B, and the extra events datasets) relative to our predictions are plotted against distance and magnitude in Figures 4.18 and 4.19 for the vertical and horizontal data respectively. There is no evidence for the residuals to be significantly dependent on magnitude or distance apart from at low frequencies and magnitudes: for the small-magnitude data (M_w < 3.5) the data are not really valid below 1 Hz, as this is often within the background noise. This means that the ground motion at frequencies below around 1Hz are artificially high. Our model therefore under-predicts these very low frequencies for small earthquakes.

The error on the model was estimated by assuming that the residuals are log-normally distributed. The logarithm of the distribution of the residuals is shown in Table 4.15 for different magnitude ranges.

		3 ≤ Ml < 4	4 ≤ Ml < 5	5 ≤ Ml < 6	6 ≤ Ml < 7	Average	Standard Deviation
PGA	+2σ	-0.8	-0.58	-0.52	-0.52	-0.6	0.13
	+σ	-0.19	-0.16	-0.1	-0.23	-0.17	0.05
	-σ	0.6	0.6	0.52	0.59	0.58	0.04
	-2σ	1.04	0.96	0.87	1.34	1.05	0.2
PGV	+2σ	-0.59	-0.57	-0.52	-0.66	-0.59	0.06
	+σ	-0.15	-0.24	-0.1	-0.37	-0.22	0.12
	-σ	0.52	0.51	0.52	0.39	0.48	0.06
	-2σ	0.93	0.83	0.87	0.9	0.88	0.04
0.5Hz	+2σ	-0.43	-0.67	-0.6	-0.78	-0.62	0.15
	+σ	-0.04	-0.34	-0.25	-0.42	-0.26	0.16
	-σ	0.66	0.48	0.55	0.51	0.55	0.08
	-2σ	1.15	0.89	0.96	1.01	1	0.11
1.0Hz	+2σ	-0.59	-0.75	-0.78	-0.6	-0.68	0.1
	+σ	-0.18	-0.39	-0.41	-0.28	-0.32	0.11
	-σ	0.51	0.42	0.56	0.51	0.5	0.06
	-2σ	0.97	0.79	1.04	1.15	0.99	0.15
5.0Hz	+2σ	-0.84	-0.65	-0.82	-0.73	-0.76	0.09
	+σ	-0.22	-0.26	-0.33	-0.42	-0.31	0.09
	-σ	0.54	0.55	0.55	0.54	0.54	0.01
	-2σ	1.02	0.9	0.99	1.43	1.09	0.24
10.0Hz	+2σ	-0.93	-0.59	-0.6	-0.68	-0.7	0.16
	+σ	-0.19	-0.19	-0.25	-0.35	-0.24	0.07
	-σ	0.58	0.58	0.47	0.58	0.55	0.06
	-2σ	1.01	0.97	0.87	1.4	1.06	0.24

Table 4.15: Data variance over all distances for different magnitude ranges for the horizontal data. ±σ represents the 68.3% confidence interval; ±2σ represents the 95% confidence interval. Units are in log(data/model).

The different magnitude ranges show little variation in the residual distribution for PGA, PGV and the PGA, 5Hz and 10Hz response with around ±0.1 standard deviation in the log residuals over different magnitudes. For large magnitude earthquakes, above M6, we have comparatively fewer data, so any offset in the residuals in Figure 4.19 may be not representative of the real distribution. The average standard deviation, over all distances and magnitudes is shown in Table 4.16.

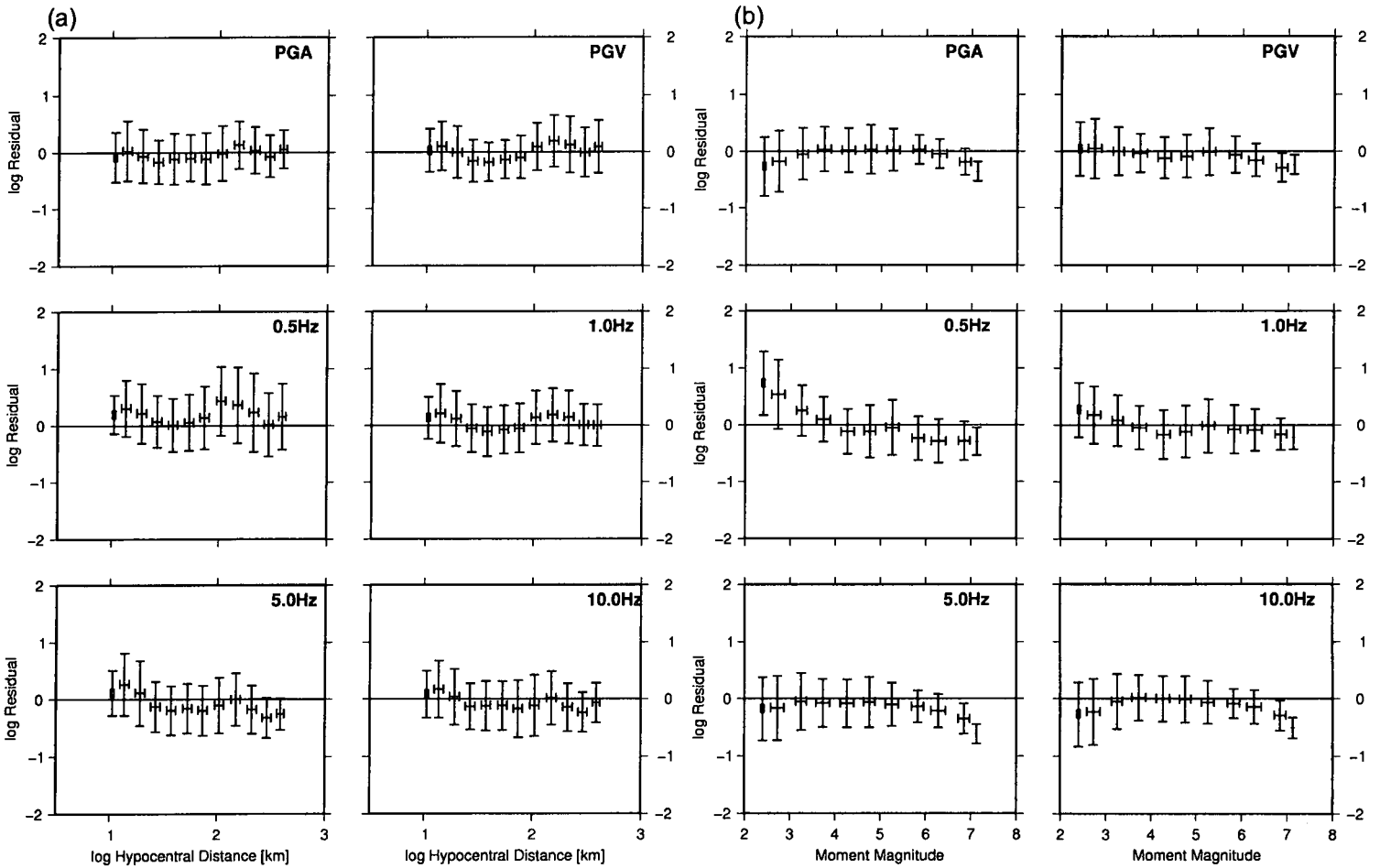

Figure 4.18: Plot of fractional residual (data/model) for all vertical data used in the study including extra large magnitude (same region) events against (a) hypocentral distance, and (b) moment magnitude. Solid line indicates no difference and error bars show one standard deviation.

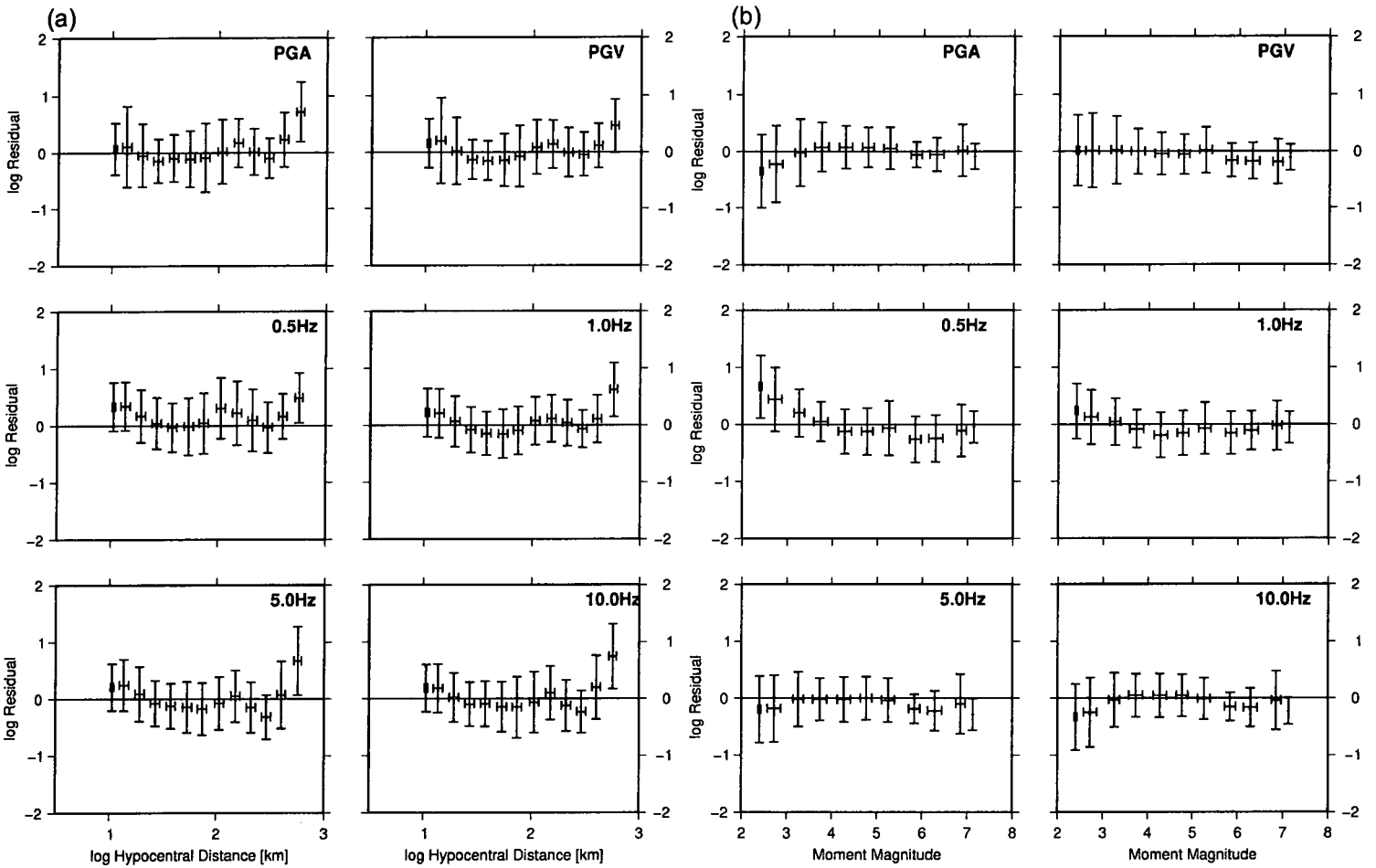

Figure 4.19: Plot of fractional residual (data/model) for all horizontal data used in the study including extra large magnitude (same region) events against (a) hypocentral distance, and (b) moment magnitude. Solid line indicates no difference and error bars show one standard deviation.

Ordinate	Average σ
PGA	0.37
PGV	0.35
0.5Hz	0.40
1.0Hz	0.41
5.0Hz	0.43
10.0Hz	0.40

Table 4.16: Average sigma value over all distances and magnitudes for each spectral ordinate.

In a recent study Cotton et al. (in press) found that there was evidence for magnitude dependent decay with distance, particularly for the intermediate frequencies of the response spectrum. They found that the ground-motion from smaller earthquakes decayed more rapidly than from large earthquakes. This was attributed to constructive interference from extended sources. Their data was from the same network as used in this study (Kik-Net). We did not consider this effect. Whilst there is no direct evidence for the requirement of this feature (for example, in the residuals plotted in Figures 4.18 and 4.19), the inclusion of such a feature should be tested in future work as it may reduce the standard deviation of residuals.

4.9 Alternative Stress Drop Model

Whilst we showed that the variable stress drop model was able to adequately model both small- and large-magnitude data, we also acknowledge that in reality the stress drop may be constant throughout the entire magnitude range. A κ filter may be instead be responsible for the perceived decrease in stress drop at low magnitudes. To test this theory the stochastic simulation is repeated as before. For the stress drop we use the constant stress drop of the large-magnitude data: 2.4MPa. In addition a site attenuation filter $\exp(-0.04\pi f)$ is applied. All other parameters remain the same. Figure 4.20 shows the residuals using this model. Some extra deviation of the mean residuals is observed relative to the residuals using the variable stress drop model (Figure 4.18). However, the differences are very small suggesting that the use of a constant stress drop in addition to a relatively strong κ filter is also suitable for modelling the observed response spectra. Further work is required however, in order to ascertain how this information could be extracted from the small-magnitude data alone. As presently the constant stress drop must be obtained from large-magnitude data.

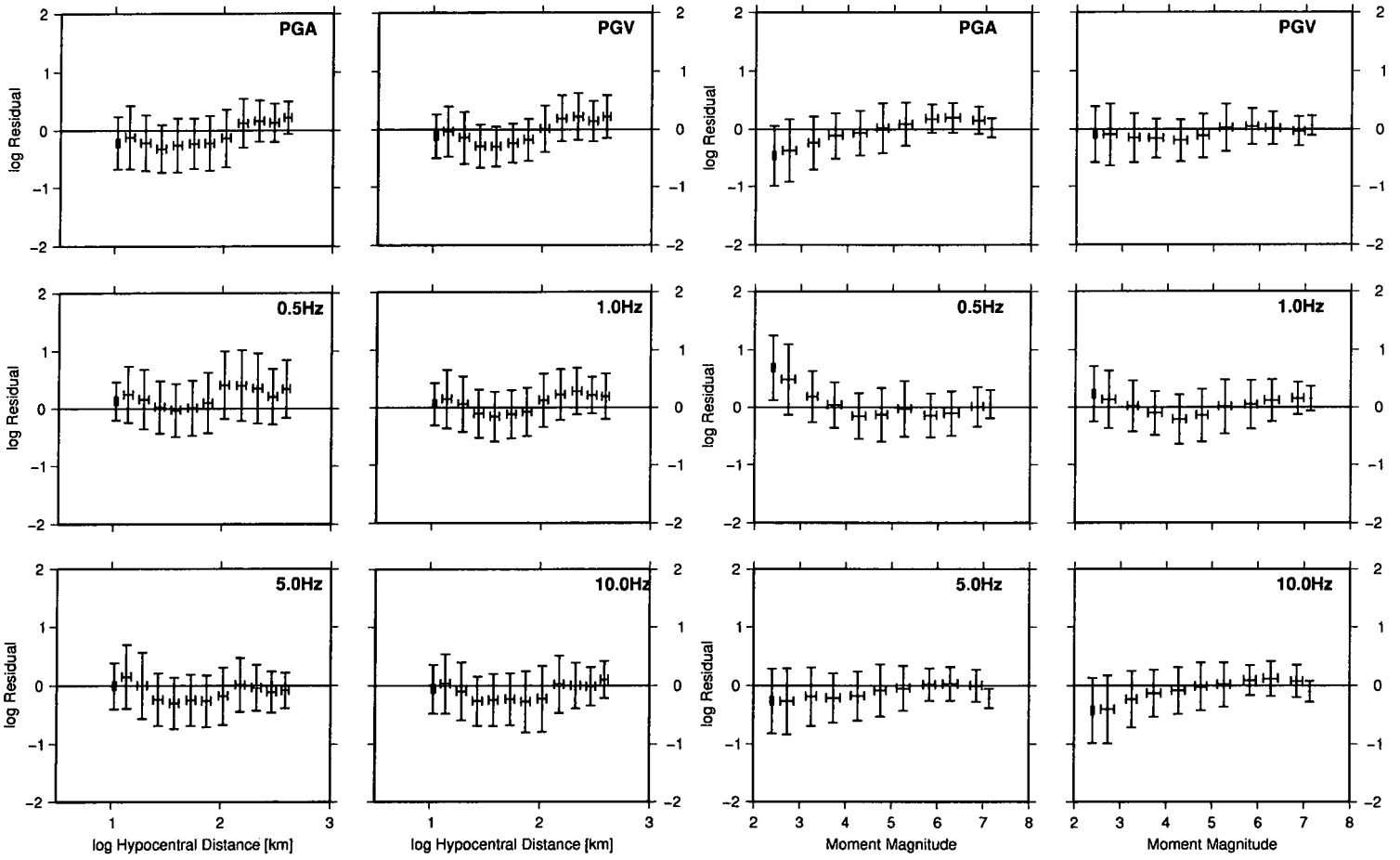

Figure 4.20: Plot of fractional residual (data/model) for the constant stress drop, κ filter (exp(-0.04πf)) model for all vertical data used in the study against (a) hypocentral distance, and (b) moment magnitude. Solid line indicates no difference and error bars show one standard deviation.

4.10 Discussion and Comparison with Other Studies

We find that the theoretical predictions of response spectral ordinates match fairly well with empirical observations. To quantify the ability of our model to predict strong ground motion we compare it with three other attenuation relations derived for the Japan region. In the text we may refer to these relations as model A: this study; model B: Pousse et al. (2006); model C: Nishimura and Horike (2003) and Model D: Horike and Nishimura (2004). Model B is derived from Japanese KiK-Net data using a model initially developed by Sabetta and Pugliese (1996) which is a "time-domain accelerogram based on the assumption that the phase is random and that the time envelope can be represented by the so-called average instantaneous power". For model B, Pousse et al. (2006) find that:

$$\log PGA = 0.4346M_W - 0.002459\,R - \log(R) + Sj \quad (4.23)$$

Models C and D are derived for Japanese K-Net (Kyoshin Network) data from the use of statistical regression methods to find optimum fitting functions to PGA and PGV data. For model C Nishimura and Horike (2003) obtain:

$$\log PGA = -1.579 + 0.739M_{JMA} + 0.022D - \log(R + 0.0006.100.69M_{JMA}) - 0.0025R.D0.263 + Sj \quad (4.24)$$

And for Model D:

$$\log PGV = -3.55 + 0.923M_{JMA} + 0.011D - \log(R + 100.992M_{JMA} - 5) - 0.0055R + Sj \quad (4.25)$$

In all these models the the site term, Sj, is set to 0 as we are predicting borehole PGA, which assumes that the site term is due only to the uppermost 100m or so. Figure 4.21 shows these relations plotted together, along with the model from this study, for an event located at 10km in depth. We convert from the average PGA given by the mean of the two horizontal components to the definition used in this study: the vector sum of the PGA from the individual horizontal components. This is achieved by multiplying by a factor of 1.4.

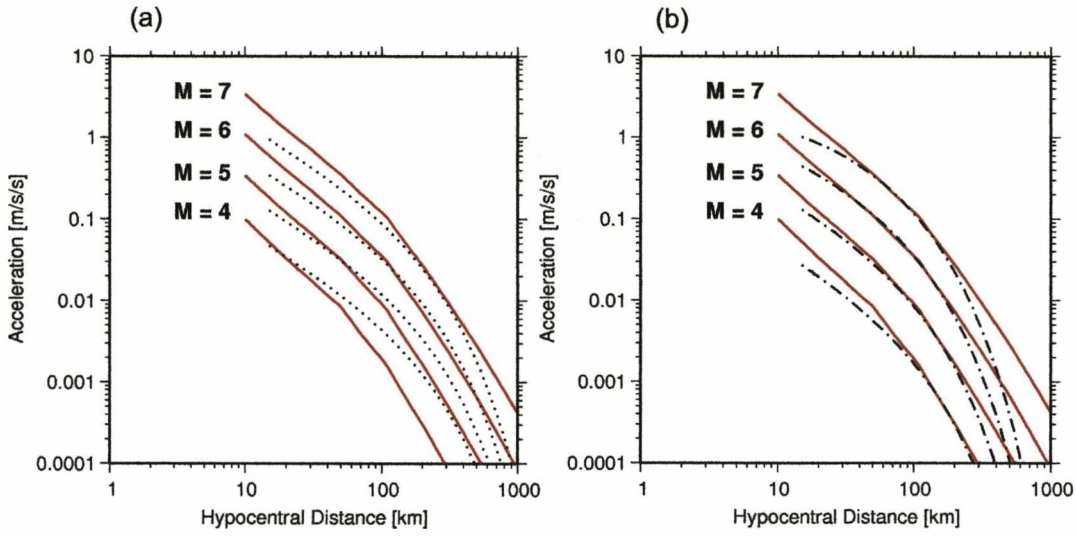

Figure 4.21: Plot of PGA values for earthquakes of moment magnitude 4.0 to 7.0 with different predictive equations. In the plots the solid lines show the relationship derived in this study. In addition, the models from (a) Pousse et al. (2006); and (b) Nishimura and Horike (2003) are plotted with dashed or dotted lines.

Figure 4.21 shows that the form of the the models is mostly similar. Model C and the model A (this study) are very similar over all magnitudes and distances of less than 100km from the source. However, for magnitudes greater than around 6 the decay of PGA beyond 100km is far higher in model C. Model B has a very similar shape of decay to model A, but shows less variance in PGA with varying magnitudes at large (100km+) distances. In order to test the ability of the models to predict empirical PGA we plot the residuals of the model compared to the data used in this study in Figure 4.22.

Pousse et al. (2006) concluded that the error on their log-PGA estimate was equal to 0.30. This is smaller than the value we obtain for our model: 0.37. If we look at only the data from magnitudes greater than 5 we obtain a value of 0.36; so the increased error is not simply due to the small-magnitude data. It is possible that the inclusion of site response values would decrease this error, but unless those site terms are linked to geo-technical information their use is limited when predicting ground motion away from the recording site. The residuals in Figure 4.22 show that the two alternative models lead to a dependence of the residual on hypocentral distance: the closest and farthest data are under-predicted relative to the mean distances. Model C gives residuals that are mostly independent of magnitude. However, for model B the residuals are dependent on magnitude for magnitudes less than around 4: Below this the model of Pousse et al. (2006) greatly over-predicts the PGA. Clearly this problem is also apparent when comparing model B with the dataset in this study. However, the ability of the model proposed in this study and the model of Nishimura and Horike (2003) to predict PGA across the complete magnitude range suggests that it is not an inherent shortcoming of predictive equations.

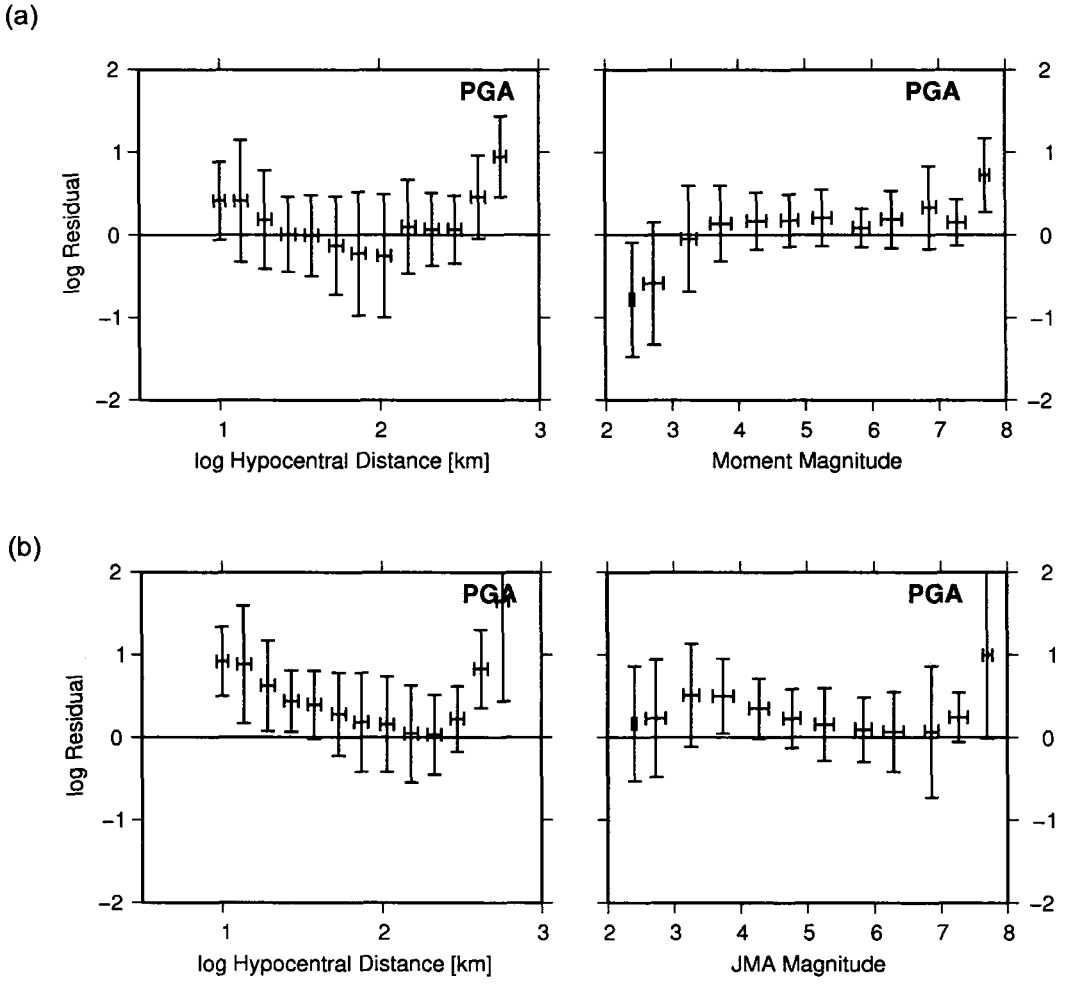

Figure 4.22: PGA residuals are plotted for all the data in the study (as in Figure 4.19) using the models of (a) Pousse et al. (2006) and (b) Nishimura and Horike (2003).

In addition to PGA, the model of Horike and Nishimura (2004) (Model D) predicts PGV. Figure 4.23 compares this with the relation from this study for an event at 10km in depth. The residuals for this model are also plotted in Figure 4.24. The comparison between PGV from model A and D is similar to that of PGA: closer than 100km the models are comparable, beyond this magnitudes 5 and above give much lower PGV using model D than from model A.

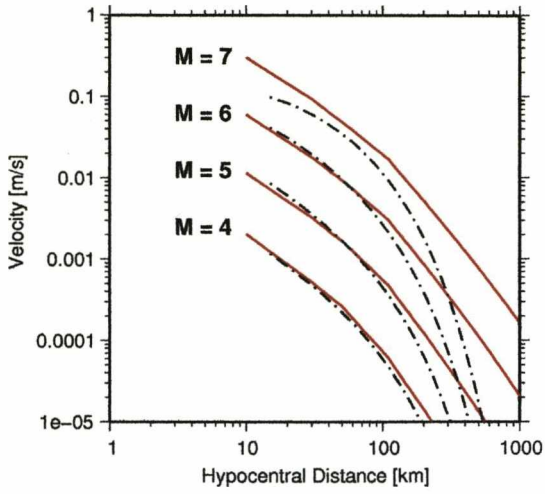

Figure 4.23: Plot of PGV values for earthquakes of moment magnitude 4.0 to 7.0 for the model in this study (solid lines) and the model of Horike and Nishimura (2004) (dashed line).

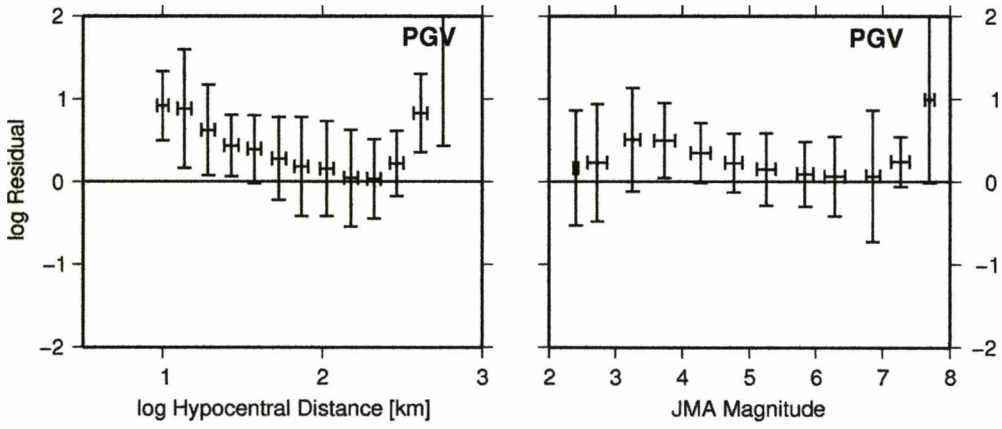

Figure 4.24: PGV residuals are plotted for all the data in the study using the model of Nishimura and Horike (2003).

The PGV residuals from model D are mostly independent of magnitude, but the model greatly under-predicts PGV at distances greater than 100km. Overall the three alternative models do a fair job of predicting PGA or PGV. There is a common feature of under-predicting very distant (over 100km) and very near (~10km) observations. The relationship derived in this study is certainly comparable to the relationships from other studies in Japan which require the use of large-magnitude data. Our study did not require the large-magnitude data, we used it simply as a tool for comparison.

4.11 Conclusions

The correlation of the two independent large- and small-magnitude datasets shows that it is possible to model attenuation parameters which are valid through a wide magnitude range using only small-magnitude data. We find that in order to model both the small- and large-magnitude datasets the stress drop must increase proportional to $M_W^{0.24}$, with a value of

0.91MPa at magnitude 3. Alternatively the stress drop may be constant, and the decrease in stress drop for smaller magnitudes is an artifact of a static corner-frequency low-pass site filter. Apparent geometrical spreading is best described by a multiple-segment model: with an increasing rate of decay with distance. We also see that the site attenuation (kappa) is highly spatially variable. As such the direct comparison of this parameter between small- and large-magnitude datasets was difficult due to different instrument locations. There seems to be an inverse correlation between the site attenuation (kappa) and the frequency independent site amplification. This suggests that the cause of the two is somehow coupled, for example soft rocks, with partially fluid filled pores, which would attenuate the signal, with a soft soil or clay cover which would amplify the signal over a broad frequency range, and vice versa for harder igneous rocks with little soil cover. Moment magnitudes follow a polynomial trend when plotted against M_{JMA}, which is very similar to that found in Europe.

After comparing predicted and empirical measures of ground motion we found that there was no significant dependence of the residual on the magnitude or distance from the source. The success of this methodology suggests that the attenuation parameters derived earlier are valid for both weak- and strong-ground motion. Any differences, if significant in terms of the response spectrum, are cancelled out through the associated covariance matrix; this does not then significantly effect the upscaling of the ground motion prediction through stochastic simulation. We tested a homogeneous Q model (i.e., as opposed to a layered structure) in the simulations, however this led to large residuals between the model and data, particularly at low hypocentral distances where energy spends longer in the upper crust. The layered Q model is a necessity in order to account for the strong attenuation in the upper crust.

In conclusion we found that we are able to use vertical recordings from a small-magnitude dataset to simulate strong ground motion which can successfully predict the response even at magnitudes of 6.5 and more with average log space errors of around ±0.36 for PGA and PGV and around ±0.40 for the response spectrum. These relations were comparable to other relationships which used large-magnitude data from the study area. Even though we obtained site specific responses we neglected them in the stochastic simulations. By taking into account this response (although not as significant as for surface located stations), we may be able to further reduce the error associated with this prediction. However, more work is required in order to fully understand the site specific contribution to the ground motion. Further work is also required to assess the effect of the near surface to the upscaling, where it is suggested that there may a dependence on magnitude to the response of soil layers (e.g., Hartzell et al., 2004). Furthermore, the predictive equations are limited to prediction of the far-field (10km+) ground motion since the method relies on the simplification of the earthquake source by Brune (1970, 1971). This is ultimately a limitation of the available small-magnitude data: only dense recording of near-field ground motion would allow us to model this phenomena. In terms of generalising the method to other regions more testing is required. The application of this method to other areas of high seismic activity is required to

test if it is repeatable. In addition further analysis of the results of this study will be beneficial, for instance the analysis of the near-field and the near-surface effects and their implication on the prediction of the response spectrum. The quality of the small-magnitude dataset in this case study was excellent, in fact only five weeks of data were required. However, in regions of low seismic activity data may not be of such quality. The use of such data should be carefully analysed, for instance using bootstrap testing, to test for its influence on any results. The results and conclusions presented here are clearly only directly applicable to this particular region and dataset. However, rather than addressing all the limitations of the method to numerous different datasets and the quality of data required, this paper is presented as a proof-of-concept: that strong ground motion can be estimated using a small-magnitude database.

4.12 Acknowledgments

This project is funded by the Natural Environment Research Council and the British Geological Survey. The data were made available by the National Institute for Earth Science and Disaster Prevention, Japan. The figures in this paper were made using GMT 4 (generic mapping tools) software written by Wessel and Smith (1991). Parts of the code used in creating the inversion tools were provided by Pozo (2004) and Press et al. (1997).

Chapter 5

Attenuation relations for Great Britain derived from local micro-seismicity

5.1 Introduction

Attenuation relations are a key component of seismic hazard analysis. This is regardless of whether the analysis is probabilistic or deterministic. Attenuation relations are typically computed by regressing empirical earthquake data (e.g., Akkar and Bommer, 2007; Ambraseys et al, 2005, 2005b). In regions of low to moderate seismicity the derivation of these relations from local records of seismicity is limited by the range of magnitudes in the recorded earthquake catalogue. To constrain the regression of ground-motion data at high magnitudes, data from areas outside the study region are taken. In the case of the UK, attenuations were formulated in the 1980s (Principia Mechanica 1982, 1988). The Principia Mechanica study used global earthquakes from tectonically similar settings as the UK (i.e., intraplate tectonic regions). However, this practice is heavily reliant on the assumption that attenuation is not regionally variable. As shown in Chapter 3, an alternative method of formulating predictive ground motion equations or tables is to use the stochastic method along with the attenuating properties of the crust derived from micro-earthquakes. It was shown that this method was valid, and did a comparable job of predicting strong ground-motion to relations derived by statistical regression of strong-ground motion data. In Chapter 3 we derived the crustal attenuation properties, stress-drop and moment-magnitude scaling relationships for the UK. In the following chapter these parameters are used to form the basis of a stochastic simulation; computing the predicted PGA, PGV or response spectral ordinates at a range of distances for a typical earthquake. The resulting predictions of strong ground motion are then compared to relations commonly used in the UK and additionally are used to model two of the largest UK earthquakes to occur in the last 3 decades.

5.2 Method

Using the results from Chapter 3, and the stochasic simulation method of Boore (2003) we compute attenuation relations for Great Britain using a number of measures of ground-shaking. These are PGA, PGV, and the response spectral values for a 5% of critical damped simple harmonic oscillator at 0.5Hz, 1Hz, 5Hz and 10Hz.

The moment magnitudes of the data were estimated using the relationship of Grünthal and Wahlström (2004). They used a second degree polynomial to model the moment magnitude – local magnitude relation for Europe. They found that:

$$M_w = 0.67(\pm 0.11) + 0.56(\pm 0.08)\, M_L + 0.046(\pm 0.013)\, M_L^{\,2}. \quad (5.1)$$

For the magnitude range of the data available ($2 < M_w < 4.5$) this relationship follows:

$$M_w = 0.71M_L + 0.58, \quad (5.2)$$

as found in Chapter 3. For $M_L<3$ Equations 5.1 and 5.2 are almost identical. This is shown in Figure 5.1. The relationship of Grünthal and Wahlström (2004) was chosen due to its compatibility with the results from Chapter 3, but that it is also valid at higher magnitudes.

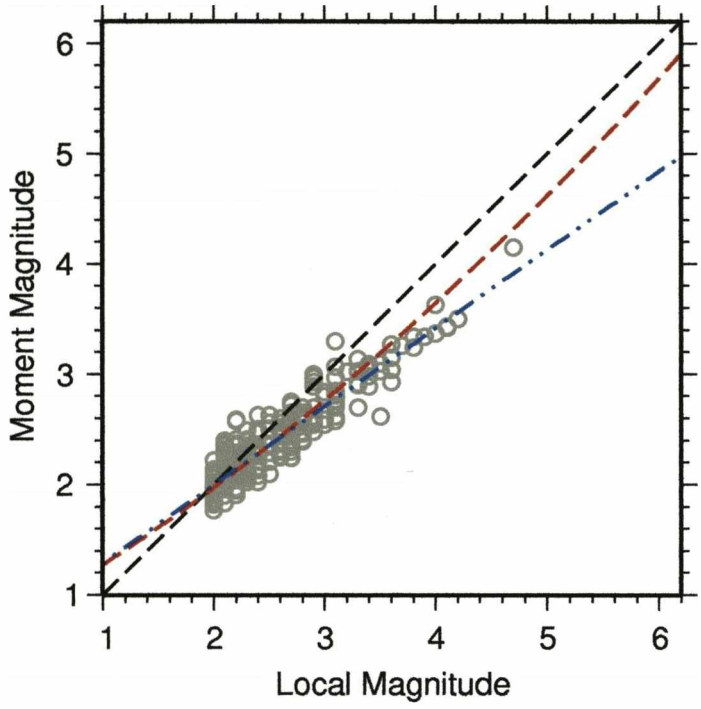

Figure 5.1: Plot of local magnitude (provided by the British Geological Survey) against computed moment magnitude from Chapter 3. The trends are a 1:1 relationship (black dashed line); a European relationship (Grünthal and Wahlström, 2004) (dashed line); and the L2 relationship given in Chapter 3 (dot-dashed line).

The duration of shaking can be defined by:

$$T_{dur} = \frac{1}{f_c} + T_{path} \quad (5.3)$$

For the path duration we visually inspect a density plot of the velocity time series over a range of hypocentral distances (Figure 5.2).

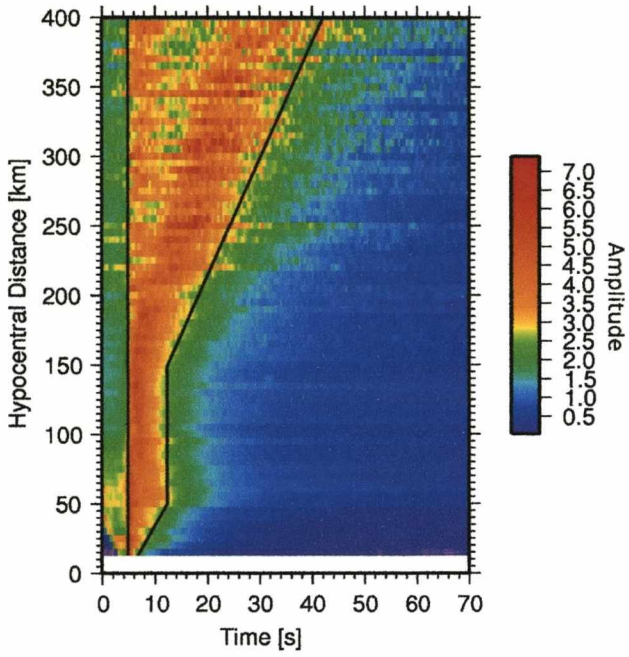

Figure 5.2: Plot of stacked RMS velocity traces over a range of hypocentral distances. The amplitude of each trace is normalised such that the peak velocity is equal to 1. Traces at equal hypocentral distances are summed to give the amplitude values given in the plot. The time axis is normalised such that the S wave onset is always at 5 seconds. The duration used for stochastic simulation is shown by the black line.

From this we determined that the average duration of shaking could be defined by the values in the following table (Table 5.1). This is also highlighted in Figure 5.2.

Distance (km)	0	60	150	400
T_{path} (s)	0	7	7	37

Table 5.1: Path duration of shaking at several distances. Linear interpolation between these points gives the function used in calculating T_{path}.

The source corner frequency, f_c, was estimated from the moment magnitude assuming the relations:

$$\Delta\sigma = \frac{7}{16} \cdot \frac{M_0}{r_0^3} \quad , \tag{5.4}$$

(Eshelby, 1957), and

$$r_0 = \frac{v_s k}{f_c} \quad , \tag{5.5}$$

Brune (1970, 1971). In Chapter 4 it was found that the stress drop may actually be constant, and the increasing stress drop with magnitude was an artefact cased by a strong site low-pass filter term. However, it is not possible to obtain this information, at present, from the the small-magnitude dataset available in the UK. Therefore the variable stress-drop relation is used, as it was found to be adequate for modelling strong-motion data in Japan (Chapter 4).The stress drop is therefore given by:

$$\Delta\sigma = \Delta\sigma_{ref} \cdot 10^{0.27(M_w - M_{ref})} \tag{5.6}$$

where $\Delta\sigma_{ref}$ = 1.1MPa at M_{ref} = 3 such that the source corner frequency is related to the moment magnitude by:

$$log(f_c) = -0.41M_w + 1.94, \tag{5.7}$$

as found in Chapter 3. The rate of apparent geometrical decay is a 3-segment function of constant exponential decay:

$$S_{ij}(r) = \begin{cases} \left(\frac{r_0}{r}\right)^{\lambda_1} & r \le r_1 \\ S(r_1) \cdot \left(\frac{r_1}{r}\right)^{\lambda_2} & r_1 \le r \le r_2 \\ \vdots & \vdots \\ S(r_n) \cdot \left(\frac{r_n}{r}\right)^{\lambda_n} & r \ge r_n \end{cases} \quad . \tag{5.8}$$

Where n=3 and the decay values, λ, are given in Table 5.2.

Distance Range	0-50 km	50-100km	100km +
Rate of Decay, λ	1.0	0.00	2.22

Table 5.2: Rate of decay values for Equation 5.8.

The layered path attenuation structure from Chapter 3 was used. In order to apply this structure to the method of Boore (2003), which uses a homogeneous Q structure, we computed the average Q value for several discrete path lengths from a source at 10km depth. An example of these are given in Table 5.3.

Path Length (km)	Q
10	900
100	1000
150	1960
200	2350
300	3000
500	3770

Table 5.3: Path Q at several hypocentral distances.

Site terms (a(f), κ, A) were not used in the simulation. This was due to the assumption in the methodology of Chapter 1 that the average κ, a(f) and A must all be negligable over all recordings.

5.3 Results

Using these parameters the programme SMSIM (Boore, online) was used to compute stochastic simulations of PGA, PGV, and the response spectra over a range of distances and magnitudes. Tables 5.4 to 5.9 summarise these predictions for a source at 10km depth. The tables use less distance nodes than in the computations used for producing the figures and residuals, but with logarithmic interpolation between the distance and magnitude nodes an estimate at any distance and magnitude is possible.

As a rule of thumb, the model given in tables 5.4 to 5.9 my be converted to the vertical plane through division by 1.5 (e.g., Newmark and Hall, 1982). Although in reality the ratio between the horizontal and vertical plane may be dependent on frequency. The performance of these predictions of ground-shaking can be tested by computing the response spectra for the records of ground motion we have. This will, however, only test their performance within the range of magnitudes available in the earthquake catalogue. Never-the-less, in a study of both weak and strong-motion records from Japan we found that we were able to predict the strong motion ground shaking using only recordings of micro-seismicity (Chapter 4). Figure 5.3 shows the residuals when comparing the vertical data and predictions. The residuals are largely independent of magnitude and distance with a mean error values given in Table 5.10.

	log PGA (m/s/s)				
log Dist. (km)	M_w = 3.0	M_w =4.0	M_w = 5.0	M_w = 6.0	M_w =7.0
1.00	0.89	1.56	2.18	2.76	3.30
1.10	0.69	1.37	2.00	2.59	3.14
1.20	0.48	1.18	1.82	2.42	2.98
1.30	0.27	0.98	1.64	2.25	2.81
1.40	0.06	0.78	1.45	2.07	2.65
1.50	-0.16	0.58	1.26	1.89	2.48
1.60	-0.38	0.38	1.07	1.71	2.31
1.70	-0.60	0.17	0.87	1.53	2.14
1.80	-0.72	0.07	0.78	1.45	2.07
1.90	-0.81	0.00	0.72	1.39	2.01
2.00	-0.90	-0.08	0.66	1.33	1.96
2.10	-1.08	-0.26	0.47	1.14	1.77
2.20	-1.26	-0.46	0.27	0.94	1.57
2.30	-1.60	-0.79	-0.06	0.63	1.27
2.40	-1.95	-1.12	-0.38	0.32	0.98
2.50	-2.27	-1.44	-0.68	0.02	0.69

Table 5.4: Table of horizontal PGA values for distances spaced logarithmically between 10 and 316.2km and moment magnitudes between 3 and 7. The earthquake source is located at 10km in depth.

	log PGV (m/s)				
log Dist. (km)	M_w = 3.0	M_w =4.0	M_w = 5.0	M_w = 6.0	M_w =7.0
1.00	-1.27	-0.40	0.42	1.19	1.93
1.10	-1.42	-0.54	0.29	1.08	1.82
1.20	-1.57	-0.68	0.16	0.95	1.71
1.30	-1.74	-0.83	0.03	0.83	1.59
1.40	-1.90	-0.98	-0.11	0.71	1.48
1.50	-2.07	-1.13	-0.25	0.58	1.36
1.60	-2.24	-1.29	-0.39	0.45	1.24
1.70	-2.41	-1.44	-0.54	0.31	1.12
1.80	-2.49	-1.50	-0.58	0.28	1.10
1.90	-2.53	-1.52	-0.60	0.27	1.09
2.00	-2.58	-1.56	-0.62	0.26	1.08
2.10	-2.78	-1.76	-0.83	0.04	0.86
2.20	-2.99	-1.99	-1.06	-0.19	0.63
2.30	-3.32	-2.30	-1.36	-0.47	0.37
2.40	-3.63	-2.60	-1.65	-0.76	0.10
2.50	-3.94	-2.90	-1.94	-1.04	-0.17

Table 5.5: Table of horizontal PGV values for distances spaced logarithmically between 10 and 316.2km and moment magnitudes between 3 and 7. The earthquake source is located at 10km in depth.

	log 0.5 Hz Response (5% damped) (m/s/s)				
log Dist. (km)	M_w = 3.0	M_w =4.0	M_w = 5.0	M_w = 6.0	M_w =7.0
1.00	-1.90	-0.62	0.64	1.73	2.52
1.10	-2.03	-0.73	0.54	1.63	2.42
1.20	-2.16	-0.84	0.45	1.53	2.32
1.30	-2.28	-0.95	0.35	1.43	2.22
1.40	-2.40	-1.05	0.26	1.33	2.11
1.50	-2.51	-1.15	0.15	1.23	2.01
1.60	-2.63	-1.26	0.05	1.12	1.90
1.70	-2.75	-1.37	-0.05	1.02	1.79
1.80	-2.77	-1.38	-0.06	1.01	1.79
1.90	-2.78	-1.39	-0.07	1.01	1.78
2.00	-2.79	-1.40	-0.07	1.00	1.78
2.10	-3.01	-1.61	-0.29	0.78	1.56
2.20	-3.23	-1.84	-0.51	0.56	1.33
2.30	-3.49	-2.09	-0.77	0.30	1.08
2.40	-3.76	-2.36	-1.02	0.05	0.83
2.50	-4.03	-2.62	-1.29	-0.21	0.58

Table 5.6: Table of horizontal 0.5Hz response spectral values for distances spaced logarithmically between 10 and 316.2km and moment magnitudes between 3 and 7. The earthquake source is located at 10km in depth.

	log 1.0 Hz Response (5% damped) (m/s/s)				
log Dist. (km)	M_w = 3.0	M_w =4.0	M_w = 5.0	M_w = 6.0	M_w =7.0
1.00	-1.20	0.12	1.28	2.13	2.79
1.10	-1.31	0.02	1.18	2.03	2.69
1.20	-1.41	-0.08	1.08	1.92	2.59
1.30	-1.52	-0.18	0.97	1.81	2.48
1.40	-1.63	-0.29	0.87	1.71	2.37
1.50	-1.75	-0.40	0.76	1.59	2.26
1.60	-1.87	-0.52	0.64	1.48	2.15
1.70	-2.00	-0.63	0.52	1.37	2.04
1.80	-2.02	-0.65	0.51	1.35	2.02
1.90	-2.03	-0.66	0.50	1.34	2.02
2.00	-2.05	-0.67	0.49	1.33	2.01
2.10	-2.26	-0.89	0.27	1.11	1.79
2.20	-2.49	-1.12	0.05	0.89	1.57
2.30	-2.77	-1.39	-0.23	0.62	1.31
2.40	-3.05	-1.66	-0.50	0.35	1.04
2.50	-3.32	-1.94	-0.77	0.08	0.78

Table 5.7: Table of horizontal 1Hz response spectral values for distances spaced logarithmically between 10 and 316.2km and moment magnitudes between 3 and 7. The earthquake source is located at 10km in depth.

	log 5.0 Hz Response (5% damped) (m/s/s)				
log Dist. (km)	**M_w = 3.0**	**M_w =4.0**	**M_w = 5.0**	**M_w = 6.0**	**M_w =7.0**
1.00	0.45	1.38	2.08	2.70	3.27
1.10	0.33	1.26	1.97	2.59	3.16
1.20	0.20	1.13	1.84	2.47	3.05
1.30	0.07	1.00	1.72	2.35	2.93
1.40	-0.07	0.87	1.59	2.22	2.81
1.50	-0.21	0.73	1.45	2.09	2.69
1.60	-0.36	0.58	1.31	1.96	2.56
1.70	-0.51	0.43	1.16	1.82	2.43
1.80	-0.56	0.38	1.11	1.77	2.39
1.90	-0.59	0.35	1.08	1.74	2.36
2.00	-0.63	0.32	1.05	1.71	2.33
2.10	-0.84	0.11	0.84	1.50	2.12
2.20	-1.05	-0.11	0.63	1.29	1.91
2.30	-1.36	-0.42	0.32	1.00	1.63
2.40	-1.67	-0.72	0.02	0.70	1.35
2.50	-1.97	-1.02	-0.28	0.41	1.06

Table 5.8: Table of horizontal 5Hz response spectral values for distances spaced logarithmically between 10 and 316.2km and moment magnitudes between 3 and 7. The earthquake source is located at 10km in depth.

	log 10.0 Hz Response (5% damped) (m/s/s)				
log Dist. (km)	**M_w = 3.0**	**M_w =4.0**	**M_w = 5.0**	**M_w = 6.0**	**M_w =7.0**
1.00	0.88	1.64	2.29	2.88	3.44
1.10	0.75	1.50	2.16	2.76	3.32
1.20	0.61	1.37	2.03	2.63	3.20
1.30	0.46	1.22	1.89	2.50	3.08
1.40	0.31	1.07	1.74	2.37	2.95
1.50	0.15	0.92	1.59	2.22	2.81
1.60	-0.02	0.75	1.43	2.07	2.66
1.70	-0.20	0.58	1.26	1.91	2.51
1.80	-0.28	0.50	1.19	1.84	2.45
1.90	-0.33	0.44	1.13	1.78	2.40
2.00	-0.41	0.38	1.07	1.72	2.33
2.10	-0.60	0.18	0.87	1.52	2.13
2.20	-0.79	-0.01	0.68	1.33	1.95
2.30	-1.12	-0.34	0.36	1.02	1.66
2.40	-1.45	-0.66	0.04	0.71	1.36
2.50	-1.76	-0.98	-0.27	0.40	1.06

Table 5.9: Table of horizontal 10Hz response spectral values for distances spaced logarithmically between 10 and 316.2km and moment magnitudes between 3 and 7. The earthquake source is located at 10km in depth.

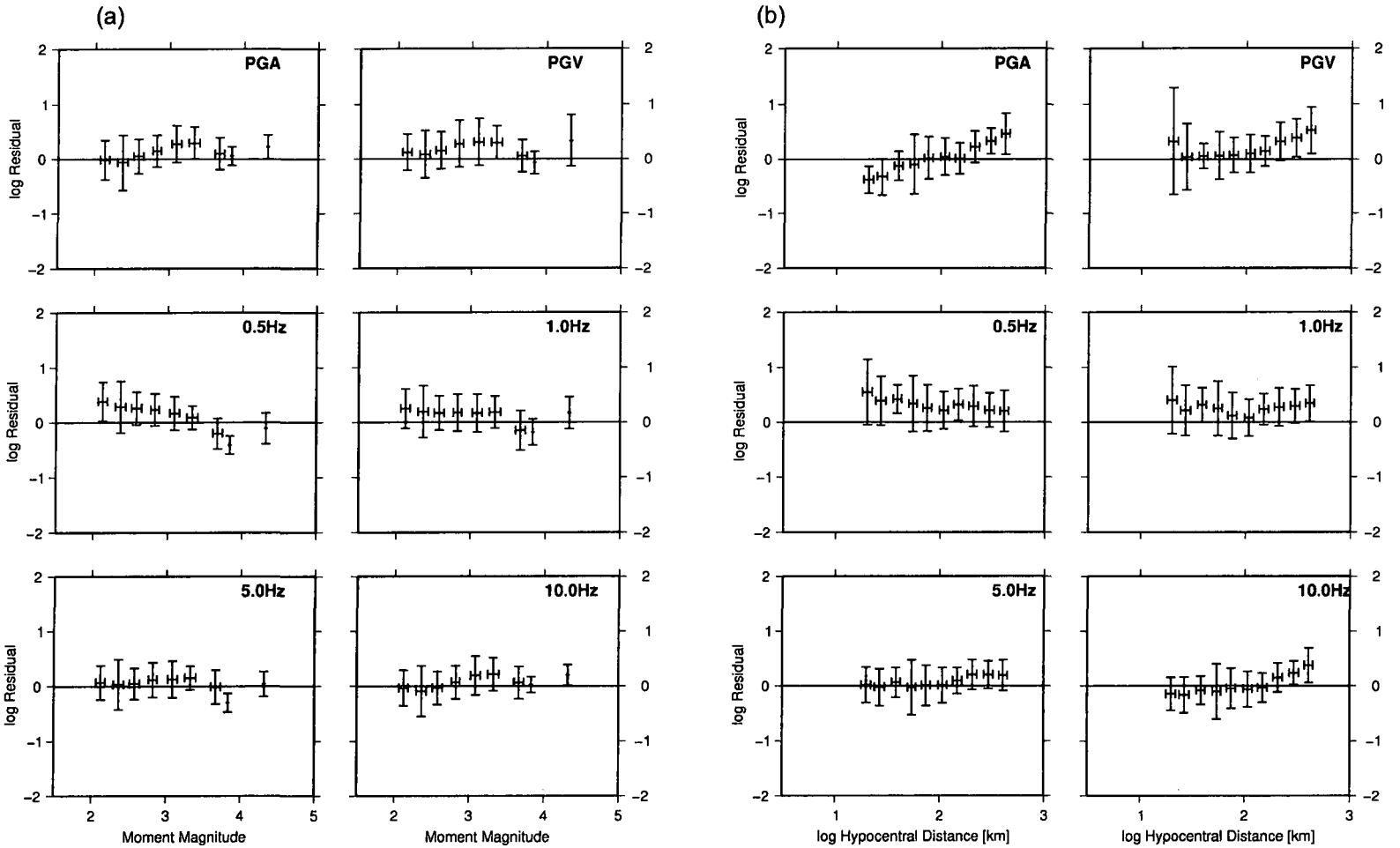

Figure 5.3: Plot of the fractional residuals between the data and the predicted model for different response spectral ordinates. Left: residuals plotted against moment magnitude. Right: residuals plotted against hypocentral distance.

PGA	-2σ	-0.83
	-σ	-0.46
	+σ	0.19
	+2σ	0.59
PGV	-2σ	-0.59
	-σ	-0.31
	+σ	0.23
	+2σ	0.72
0.5Hz	-2σ	-0.53
	-σ	-0.18
	+σ	0.36
	+2σ	0.88
1.0Hz	-2σ	-0.58
	-σ	-0.29
	+σ	0.3
	+2σ	0.73
5Hz	-2σ	-0.72
	-σ	-0.39
	+σ	0.16
	+2σ	0.45
10Hz	-2σ	-0.81
	-σ	-0.48
	+σ	0.12
	+2σ	0.46

Table 5.10: Mean error values for each log spectral response ordinate over the complete magnitude and distance range. ±σ indicates the 68.3 confidence interval and ±2σ indicates the 95% confidence interval. The error is given in log units.

5.3.1 The M_L = 4.7 Dudley Earthquake, 2002.

One of the largest earthquakes to hit the UK in recent times was that occurred in Dudley on the 22nd of September 2002. According to the British Geological Survey the earthquake had a magnitude of M_L = 4.7 and the source was located at 52.53°N 2.16°W at a depth of 14km. The event was modelled by Baptie et al. (2005) who found that M_w = 4.1±0.12 and a stress drop of 3.1±1.6MPa. From Equation 5.1 we estimated that M_w = 4.3. Using the predictive attenuation relations given in Tables 5.4 to 5.9 we computed PGA, PGV and the response spectrum for the Dudley earthquake. In order to compare the recordings of this event with this model we filtered the vertical records between 1 and 25Hz using a 3 pole butterworth bandpass filter to remove data outside the passband of the recording instrument. These data are plotted against the vertical stochastic model for a M_w = 4.3 event at a depth of 10km (Figure 5.4).

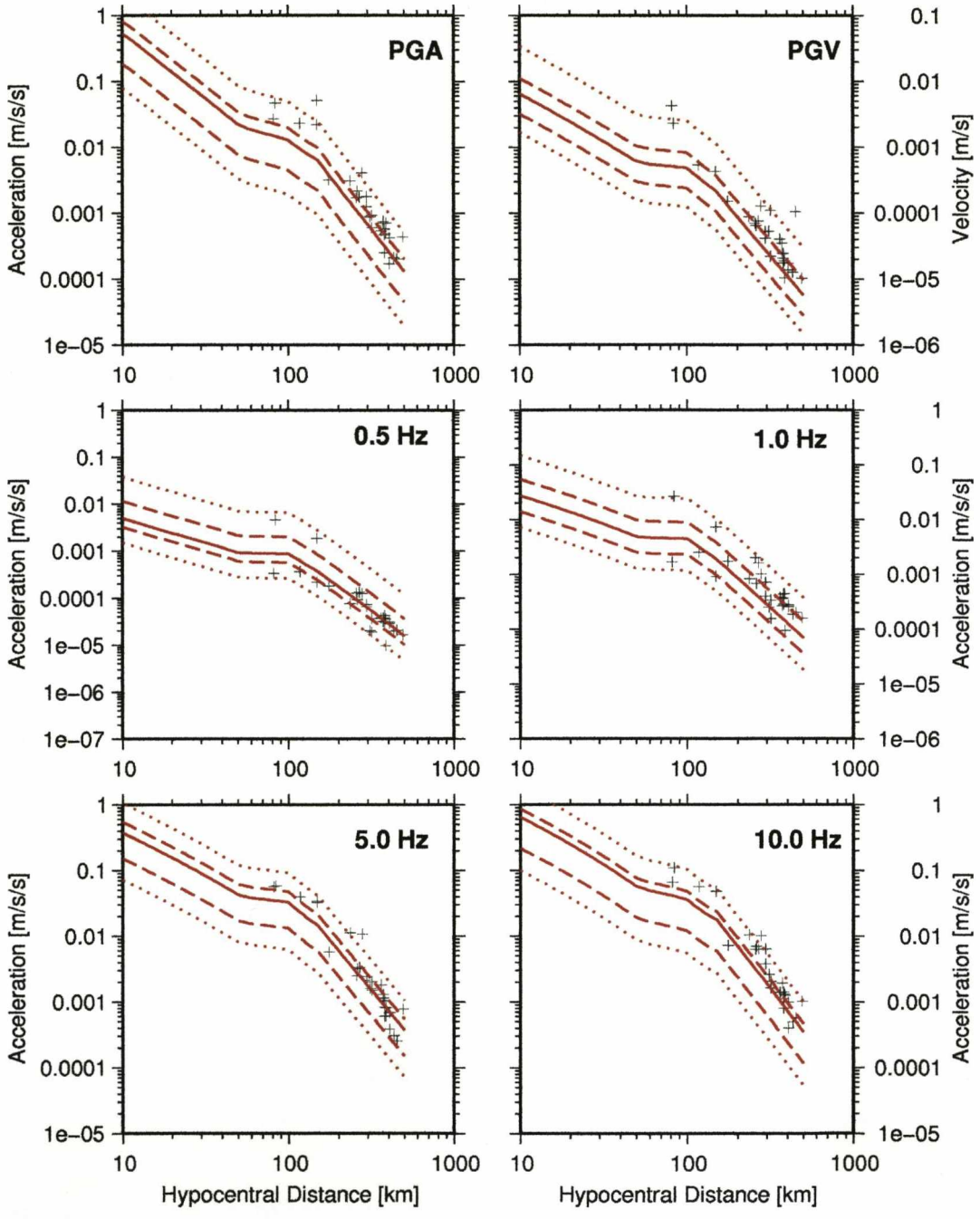

Figure 5.4: Predictions of vertical response spectral values for the 22nd of September, 2002, Dudley M_L=4.7 (M_w=4.3) earthquake. The solid line indicates the stochastic model. The dashed line indicates ±σ, the dotted line indicates ±2σ.

The model predictions are comparable to the empirical observations of the earthquake. The confidence intervals provided in Table 5.10 and shown by the dashed and dotted lines in Figure 5.4 are also a good indication of the scatter of the observations around the model. In this instance it is not necessary that the model lies at the median of empirical observations of a single earthquake. This is due to the fact that we did not consider aspects such as rupture

mechanics, style of faulting, etc.: the scatter around the median model is therefore somewhat contributed to by intra-event variability. Further work should be undertaken to reduce this aspect of model variability. Instead a measure of fit should instead be taken from the residuals of many earthquakes. Of course in this instance we do not have many earthquakes of reasonable magnitude, so this is tricky.

5.3.2 The M_L = 5.2 Market Rasen Earthquake, 2008.

The Market Rasen earthquake of the 27th of February 2008 is the largest onshore UK earthquake since 1984. According to the British Geological Survey the earthquake had a magnitude of M_L = 5.2 and its epicentre was located at 53.39°N 0.35°W. The earthquake caused structural damage in the vicinity of the epicentre and was categorised as an intensity 6 on the European macroseismic scale EMS-98 (European Seismological Commission, 1998). In synopsis, an intensity 6 on the EMS-98 implies 'some damage': "...felt by most indoors and by many outdoors. Many people in buildings are frightened and run outdoors. Small objects fall. Slight damage to many ordinary buildings e.g.; fine cracks in plaster and small pieces of plaster fall." Using a moment tensor inversion the European-Mediterranean Seismological Center computed that M_w = 4.6 and the British Geological Survey computed that M_w = 4.4. Equation 5.1 estimates that M_w = 4.8. This is slightly higher than both computed M_w. Never-the-less we use this value to remain consistent with the method adopted. The vertical and horizontal response spectra are computed for the data after filtering between 1 and 25Hz to remove background noise (e.g., background micro-seismicity). These data are compared to the stochastic model in Figure 5.5. We again observe that the model adequately predicts ground motion across a variety of measures (PGA, PGV, 0.5Hz, etc.). Adopting a more deterministic approach, and using M_w = 4.6, as found by the European-Mediterranean Seismological Centre would have made the prediction slightly worse, as the observations of ground-motion are already at the higher end of our confidence limits. Additionally, in this case there are a few records that fall outside the 95% confidence interval. While the values are not unreasonable, and have been carefully verified through visual inspection of the accelerograms, some data may not be reliable due to the instrument response applied (the instruments are new and not used elsewhere in this study). However, due to the nature of the event it is interesting to include it in this analysis. Additionally, as we did not consider intra-event variability in the formation of our model: the median model represents the response to the average earthquake. Of course, each earthquake may have slightly different characteristics that cause slightly higher or lower ground-motion response. This is represented by the standard deviation included in the model. It is perfectly acceptable, for instance, for all records of a particular earthquake to fall in the upper confidence interval, but it is also just as likely for records from another earthquake to fall in only the lower interval. This does suggest, that a reduction in the model uncertainty may be achieved through consideration of intra-event and inter-event variability.

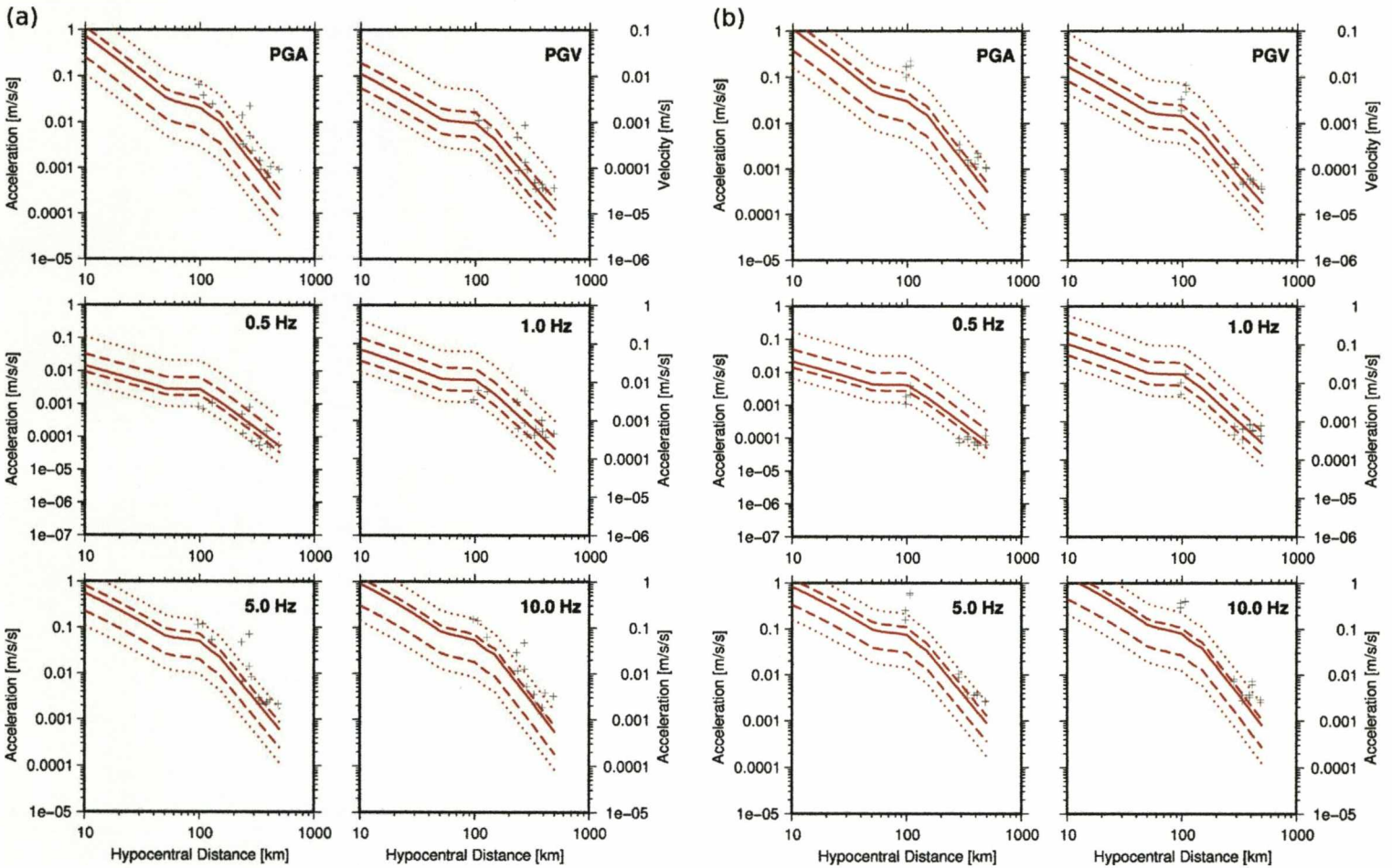

Figure 5.5: Predictions of (a) vertical and (b) mean horizontal response spectral values for the 27th of February, 2008, Market Rasen M_L=5.2 (M_w=4.6) earthquake. The solid line indicates the stochastic model. The dashed line indicates ±σ, the dotted line indicates ±2σ.

5.4 Comparison With Other Published Attenuation Relations.

While it has been shown that the attenuation relation presented above is able to successfully model UK earthquakes it is useful to compare the relation to previously published, and commonly used attenuation relations. An important aspect of this comparison is to note that although the relations are used to estimate ground motion in the UK, this is not necessarily appropriate. Nor is it the fault of the original authors that their relations are misused. It has been previously stated in this thesis that the validity of attenuation relations is really controlled by the data used in their derivation (e.g. Douglas (2003)). The use of these predictive relations outside the region for which they are derived, never mind their application to parameter ranges (such as magnitude) not included in the original study is a contentious issue. The attenuation relations compared here are used simply because they have been assumed valid in the UK, perhaps wrongly, in the past by other authors.

Ambraseys et al. (2005, 2005b) formulated predictive equations for horizontal and vertical acceleration. They are derived using data exclusively from the European and Middle East regions. The relations are an improvement on the relations of Ambraseys et al. (1996) and Ambraseys and Bommer (1991). The new relation includes the same data as the previous studies as well as newer data. This enables a "more robust estimation of the regression coefficients", Ambraseys et al. (2005). They found that:

$$\log PGA_H = 2.522 - 0.142\mathrm{M}_W + (-3.184 + 0.314\mathrm{M}_W)\log\left(\sqrt{R_e^2 + 7.6^2}\right) - 0.084 \qquad (5.9)$$

for a rock site and normal faulting. In this case R_e is the epicentral distance, or more specifically the nearest horizontal distance to the fault. Terms for soft and stiff soil sites are available, however, lack of site classifications for UK instrumentation means that these cannot be applied. Similarly, for the vertical component:

$$\log PGA_V = 0.835 + 0.083\mathrm{M}_W + (-2.489 + 0.206\mathrm{M}_W)\log\left(\sqrt{R_e^2 + 5.6^2}\right) - 0.126 \qquad (5.10)$$

In order to compare these relations with those of our study we convert the hypocentral distance to epicentral distance using:

$$R_e = \sqrt{R^2 - D^2} \qquad (5.11)$$

with R, the hypocentral distance and D the source depth.

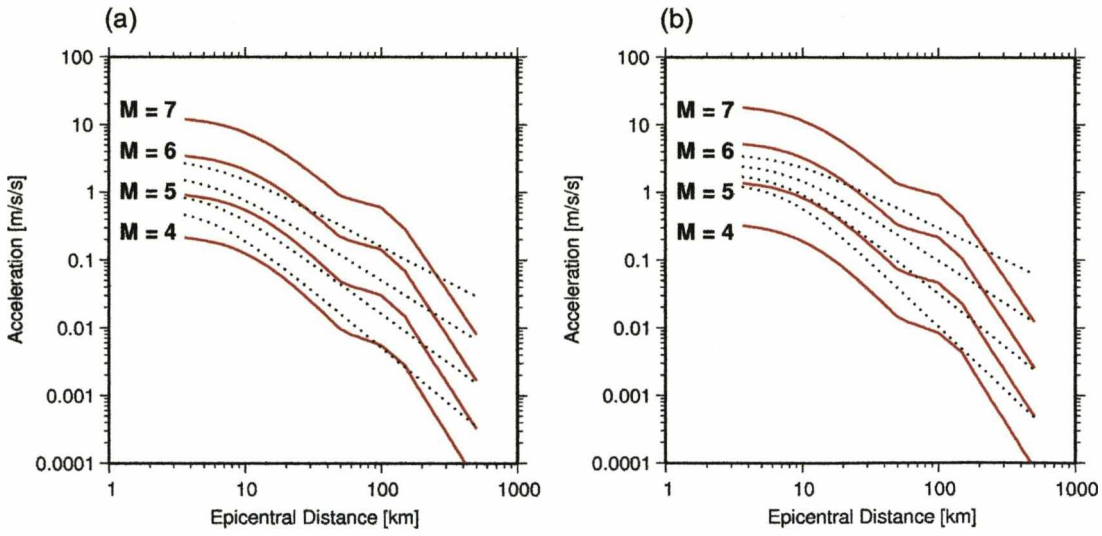

Figure 5.6: Comparison between the model in this study (red lines) and that of Ambraseys et al. (2005, 2005b) (black dotted lines). (a) Vertical PGA and (b) Horizontal PGA for a source located at 10km in depth.

Figure 5.6 shows an example of the model of Ambraseys et al. (2005, 2005b) and the model of this study using a source located at 10km in depth. Some differences exist between the models. For instance the model of Ambraseys et al. (2005, 2005b) does not exhibit the slow down in decay between around 50 and 150km. In the first 50 km from the epicentre the decay rates are similar, but the increase in PGA with increasing magnitude is far greater in the model from this study. At greater epicentral distances (100km+) the increase in PGA with magnitude is similar, however, the decay rate in Ambraseys et al. (2005, 2005b) is far lower than that in this study. As for Figure 5.3, the PGA residuals may be calculated when using the model of Ambraseys et al. (2005, 2005b). This is shown in Figure 5.7.

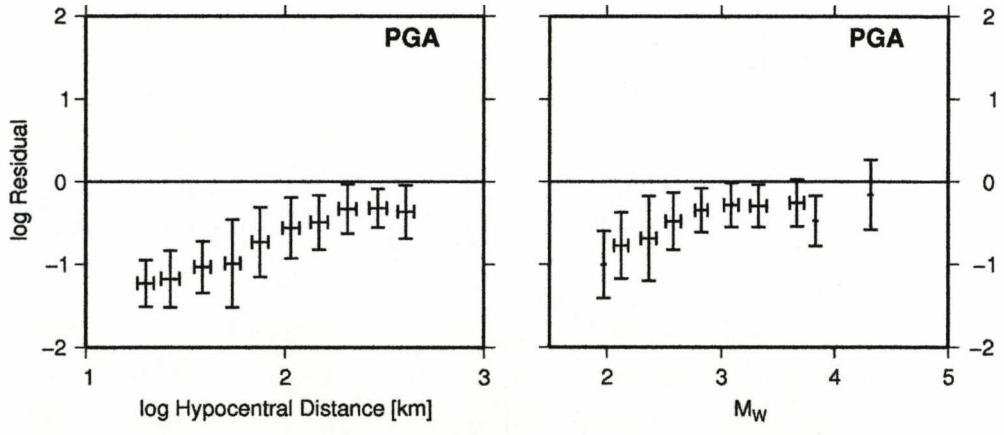

Figure 5.7: Residuals (model/data) plotted against hypocentral distance and moment magnitude when comparing the data used in this study to the model of Ambraseys et al. (2005, 2005b).

Comparing Figures 5.7 and 5.6 we can see that the model of Ambraseys et al. (2005, 2005b) over-predicts the PGA from earthquakes used in this study which are all below M_w = 4.4. Additionally a dependency on both magnitude and distance is observed in the residuals. The closest observations and the smallest earthquakes are over-estimated to a greater degree than distant observations of PGA or PGA from larger earthquakes. These observations are consistent with those of Baptie et al. (2005) who found that the predictive relations from Toro et al. (1997) and Ambraseys and Bommer (1991) over-predicted the PGA from the M_L = 4.7 Dudley Earthquake discussed previously. In fact, in Figure 5.7 the error bar located at approximately M_W = 4.4 consists of records solely from the Dudley earthquake. We observe that this event is slightly overestimated by Ambraseys et al. (2005, 2005b), but not to the same extent as smaller events. This highlights the fact that the relations, are not valid below the minimum magnitude (3.9) used in their derivation: below magnitude 3 the residuals become unacceptable. Above magnitude 3 the prediction is somewhat high, but numerous differences could account for this, including the use of different measures of distance, or magnitude.

The relation of Toro et al. (1997) is also considered valid in the UK as the data are derived from earthquakes in Central and Eastern North America. Similarly to the UK, this region is considered tectonically stable. Figure 5.8 shows a comparison between the model of Toro et al. (1997) and the model from this study. Figure 5.9 shows the residuals from the model of Toro et al. (1997) when applied to the data used in this study.

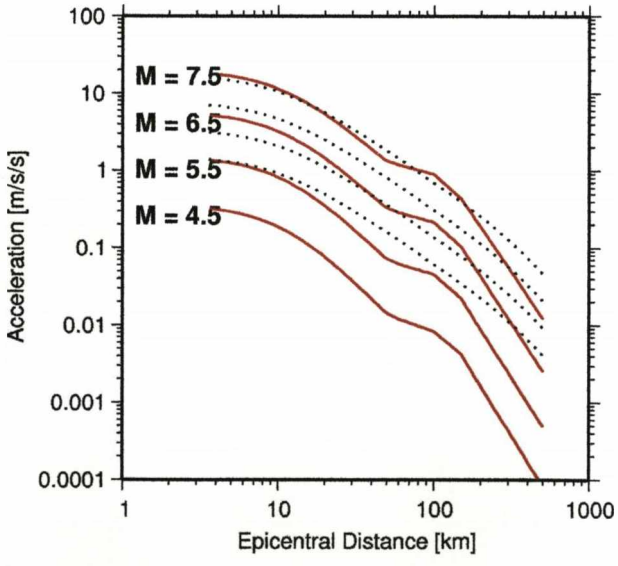

Figure 5.8: Comparison between the model in this study (red lines) and that of Toro et al. (1997) (black dotted lines). Plot of horizontal PGA for a source located at 10km in depth.

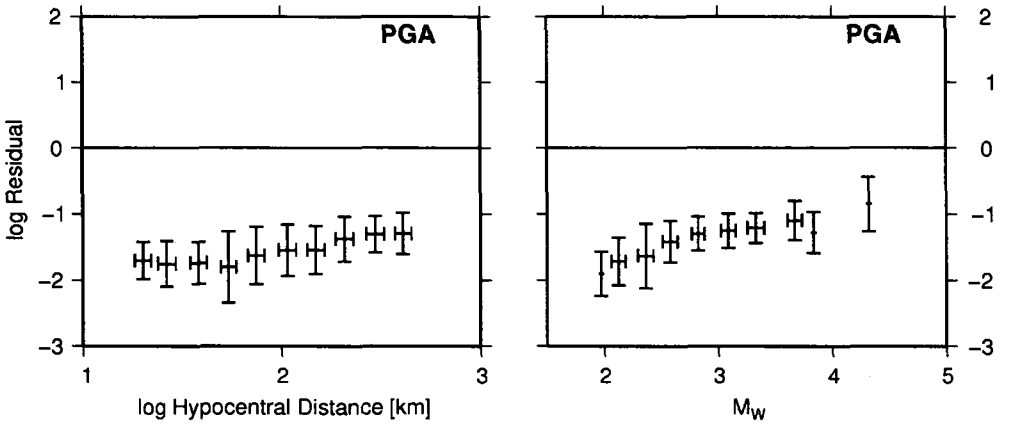

Figure 5.9: Residuals (model/data) plotted against hypocentral distance and moment magnitude when comparing the data used in this study to the model of Toro et al. (1997).

In this case PGA from the larger magnitude events (M_w>6.5) seem to match reasonably well. However, as with the model of Ambraseys et al. (2005, 2005b) the scaling with magnitude is far lower than for the model in this study. The shape of the decay in the two models are similar in the first 20km or so, after that the model of Toro et al. (1997) decays at a slower rate. As with Ambraseys et al. (2005, 2005b) we do not see the slow in decay between 50 and 150km observed in this study. The residuals (Figure 5.9) show that the model of Toro et al. (1997) greatly over-predicts the PGA of the earthquakes in this study. There seems to be a decrease in the degree of over-prediction as magnitude increases. The residuals do not seem to be dependent on distance in this case.

The aims of this comparison was not in order to compete with other relationships that have been properly calibrated for use in a different region and that are used in the valid range of parameters. Similarly, we acknowledge that the authors of the alternative relationships would also question their validity for small earthquakes. It should, however, be noted that while the damage due to ground motion from magnitudes of less than 3 is fairly insignificant, it never-the-less draws into question some important physical aspects of the models used to describe ground motion. Is there a physical process that is different during small and large earthquakes: is the model just not valid across the complete magnitude range? Or is the non-uniqueness of the regression solution to blame: the model is correct, but the solution is correct only within the parameter bounds of the data.

Chapter 6

Thesis Conclusion

6.1 Summary

In the previous chapter we used the attenuation parameters from Chapter 3 along with the stochastic method of Boore (2003) to predict response spectral ordinates (PGA, PGV and the response spectrum) of earthquakes in the UK. For the conversion of the local into moment magnitude scale we chose to use the relationship of Grünthal and Wahlström (2004) which closely follows the relationship found in Chapter 3 at low magnitudes, but is better constrained at higher magnitudes where we have little or no data. The predictions were compared to empirical observations for moment magnitudes 2 to 4.4. There were no significant dependencies of the residuals on magnitude or distance over the range available for comparison. Obviously the validity of these predictive relations beyond the magnitude range used in the derivation of the attenuation parameters is a contentious issue. However, we showed in Chapter 4 that it was possible to up-scale predictive equations that are derived from only small-magnitude data using the method detailed in Chapter 1 and throughout this thesis.

The dataset used in this study was compared to some published relations derived using strong-motion data. It was shown that these attenuation relations significantly over-predict the ground-motion due these small earthquakes. This was also found by Baptie et al. (2005) in a study of the Dudley earthquake of 2002. A difference in the scaling of magnitude between the model of this study and the compared models was apparent at short epicentral distances. This may be due to the use of a point source model and not considering finite fault effects and near-field observations.

We have found that it is possible, although not without limitations, to model moderate to large earthquakes using a database of small-magnitude events. At the start of this thesis, and throughout, it was shown that models of the Fourier spectra of earthquakes are non-unique. Covariance matrices were computed to show the variance of attenuation parameters, and also how the individual parameters behaved with respect to the others. We showed that we could improve the robustness of the solution by adopting a Q tomography approach using a variety of synthetic and bootstrap analyses. It was found that the use of a layered Q model provided more detail than a homogeneous Q model, but more stability than completely path dependent attenuation. This was most obvious in the presence of noise, as is the case with real records, where after applying the results of the Q tomography we observed a significant decrease in the error on the source corner frequency. This stability, however, was achieved at the risk of biasing the other parameters due to an incorrect initial estimation of Q. We found that the Q structure tended to be depth dependent, with higher Q towards the base of the crust. In all synthetic testing we found that the frequency dependence of Q was so strongly covariant with the other parameters that it was not possible to robustly resolve. In particular we found that the rate of apparent geometrical decay traded off strongly with the

frequency dependence of Q. It is therefore necessary to be cautious when interpreting these parameters in terms of physical processes. The source spectrum shape was well resolved in the synthetic testing suggesting that the Brune model (1970, 1971), which consistently fit the data to a better degree than alternative models, is a suitable representation of the far-field observation of the source of UK earthquakes.

In Chapter 2 synthetic seismograms were produced using finite difference modelling. This showed that the moment magnitudes computed were realistic and representative of the energy released at the source, as the moment magnitude should be. However, we also saw that the Brune stress parameter significantly underestimated the true stress drop. Synthetic seismograms were produced using the stochastic method. This allowed us to test further the estimation of the attenuation parameters, and in addition provided verification that the attenuation parameters derived would be compatible with the stochastic simulation of the response spectrum using SMSIM. We showed that as long as the average site response, defined by a(f), was negligible over many recordings of an earthquake, then the solution obtained was valid. It was also shown how the response function, a(f), would be affected if this condition was not met. For instance, an inflection in the response function may be replaced by a peak. It is important that a good range of recordings at different sensors are obtained in order to satisfy this condition.

There were some similarities between the attenuation parameters derived for the UK and Japan. For instance we found in both cases that the Q value increased with depth, although the Q of Japan was lower overall ($Q \approx 500$ to 1300 compared to $Q \approx 900$ to 5700 for the UK). The apparent geometrical decay varied with distance in both regions. However, in the UK a region of negligible decay from around 50 to 100km was observed. This was not present in the Japanese data. This is likely to be due to the differing tectonic settings of the regions: the UK is on a stable intra-plate setting, whilst Japan sits over a subduction zone. The scaling of magnitude to M_W from the M_{JMA} scale was found to follow a second degree polynomial. This was similar to a trend found for converting the European M_L scale to M_W (Grünthal and Wahlström, 2004), but with a constant positive shift in M_W. For the small magnitude range available for the UK the conversion from the M_L scale to M_W was found to follow this European relation. However, at magnitudes higher than in our dataset the least-squares relationship deviated from this European relation. As the relation of Grünthal and Wahlström, 2004 is valid at these higher magnitudes we chose to use it for this study. The site effects, both attenuation, 'κ', and amplification 'A' were found to weakly correlate in both Japanese and UK datasets. We suggested that this may be due to the correlation of site conditions, such that sites with thick soils were likely to overlie softer sedimentary rocks, whereas sites with thin soil were likely to overly harder sedimentary, igneous or metamorphic rocks. However, the analysis of these site terms was limited in this project, and further work would be beneficial in this area. The Brune stress drop was found to increase with magnitude. This

was apparent in the L2 fit of stress drops for both the UK and Japan, with a scaling factor of $M_W^{0.24}$ in Japan and $M_W^{0.27}$ for the UK derived from small-magnitude data. This scaling factor was not well resolved, with an R^2 value of around 0.4 in both datasets. However, it was found to be the better than a constant stress drop assumption derived from the small-magnitude data when compared to large magnitude events in Japan. In fact, the stress drop may be constant, and the observed decrease in the stress drop at low magnitudes may be due to a site related low-pass filter function (f_{max} (Hanks (1982) or kappa (Anderson and Hough (1984)). This site related filtering of high-frequency data results in the saturation of observed source corner frequency as it approaches the corner frequency of the filter: as magnitude decreases. Unfortunately the derivation of this combined stress drop and site filter model is not yet possible using only small-magnitude data. Further work is required to robustly deconvolve these source and site terms. As a result we proceeded to use the increasing stress drop model, as this was shown to successfully model both small- and large-magnitude events in Japan.

In Chapter 4 we used random vibration theory along with the attenuation parameters derived using small-magnitude earthquakes to simulate the response spectra of large earthquakes in central Japan. Due to the high seismicity in this region we were able to compare these simulations with empirical observations of strong ground-motion. The predictive relations performed well and we were able to show that there was no significant dependence of the residual misfit on hypocentral distance or magnitude. The predictive relations that were simulated were compared to other relations derived for Japan. The residual misfit using the other relations was worse than when using our relation, but this may be due to the derivation of these relations for surface records as opposed to borehole records. Never-the-less we found that the relations derived from small-magnitude data were at least comparable to those derived directly from strong motion data, which are not always available in regions of low seismicity.

With the knowledge that the method of formulating predictive equations to predict strong-ground motion using small-magnitude data worked, at least for our best-case scenario, we proceeded to simulate attenuation relations for the UK using the attenuation parameters previously derived (Chapter 3). Whilst there exists no records of significant strong ground-motion in the UK we were able to compare the predictions with two moderately sized earthquakes that have occurred in the UK in the past ten years: $M_L = 4.7$ and $M_L = 5.2$. It was shown that the recordings available of these events were correctly predicted by the model. We found no dependence on the distance or magnitude on the misfit residuals, although obviously this was limited to the small-magnitude recordings available.

Whilst it was shown that we were able to use a small-magnitude database to predict strong ground-motion in Japan, this study is not without its limitations. For the study of weak and

strong ground motion in Japan an 'ideal' dataset was available: numerous recordings of each event on high-quality broadband seismometers were used. The dependence of results on the quality of the dataset is something to consider, as in regions of low seismicity the uptake and funding of new recording technology is markedly less than in seismically active regions such as Japan or California. For instance, in the case of the UK's main seismic network, run by the British Geological Survey, the majority of instruments are short period 1s seismometers with high gain vertically oriented recording. As a result, these instruments tend to clip (go off-scale) for close or large magnitude events, whereas for distant events the background noise is a significant problem. In addition, in terms of obtaining the true attenuation parameters, rather than a model that is a solution to the data, the method may have shortcomings for surface recorded data. This is due to the potential for a strong site response which is not necessarily flat on average over the entire network. Whilst for the borehole recordings in Japan this assumption is reasonable, for surface stations that have their foundations in, for instance, a sedimentary basin, this may influence the perceived Brune stress parameter. Whilst this trade-off may not significantly effect the stochastic simulation in terms of the average residual over many recordings, we may encounter problems when looking to improve on the confidence limits of our predictions for particular scenarios.

As detailed in this thesis, further work is required, and is ongoing in this field of research. The separation and the understanding of source and site effects could significantly improve the prediction of ground-motion. Additionally, one limitation that is important to overcome is the applicability of the method in the near-field or for very large rupture lengths. In this case the assumption of a point source is invalid. However, there exists a limitation on how detailed our models are. This is due to the quality and quantity of data available. In the analysis of Japanese data we should be able to use a much more detailed model than for the UK due to the excellent quality recordings. We limited the detail used in order to test the applicability to regions such as the UK. Whilst it is possible to analyse in further detail, for example, the site terms and their relation to the geotechnical site conditions, it is not beneficial in terms of application to regions such as the UK where no such geotechnical information exists. In order to improve the prediction of ground motion in regions of weak seismicity we therefore require not only further work into separating the attenuation parameters, but also into the quantity and quality of the data available.

6.2 Thesis Acknowledgments

I wish to thank all my colleagues, friends and family who have helped me during my research. In particular I would like to thank my PhD supervisors Professor Andreas Rietbrock, Doctor Brian Baptie and Professor Julian Bommer, and my fiancée Janice Ward, without whom this work would have been impossible. I would also like to express my graditude towards the examiners of this Thesis, Professor Frank Scherbaum and Professor Nick Kusznir.

Bibliography

Aki, K., and P. Richards (1980). *Quantitative Seismology*: W. H. Freeman, New York, pp. 932.

Akkar, S., and J. J. Bommer (2007). Empirical prediction equations for peak ground velocity derived from strong-motion records from Europe and the Middle East, *Bull. Seismol. Soc. Am.,* **97,** 511-530.

Ambraseys N. N. and Bommer J. J. (1991). The attenuation of ground accelerations in Europe, *Earthquake Engineering and Structural Dynamics,* **20**, 1179-1202.

Ambraseys N. N., Simpson K. A., Bommer J. J. (1996). Prediction of horizontal response spectra in Europe, *Earthquake Engineering and Structural Dynamics*, **25**, 371-400.

Ambraseys, N. N., J. Douglas, S. K. Sarma, and P. M. Smit. (2005). Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: horizontal peak ground acceleration and spectral acceleration, *Bulletin of Earthquake Engineering,* **3**, 1-53.

Ambraseys, N. N., J. Douglas, S. K. Sarma, and P. M. Smit. (2005b). Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: Vertical peak ground acceleration and spectral acceleration, *Bulletin of Earthquake Engineering,* **3**, 55-73.

Anderson, J. G., and S. E. Hough (1984). A model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies, *Bull., Seismol., Soc. Am.*, **74**, 1969-1993.

Aoi, S., T. Kunugi, and H. Fujiwaras (2004). Strong-motion seismograph network operated by NIED: K-NET and KiK-net. Japan Association for Earthquake Engineering; Journal of Japan Association for Earthquake Engineering, Vol. 4, No. 3 (Special Issue), 65 – 74.

Atkinson, G., and R. Mereu (1992). The shape of ground motion attenuation curves in southeastern Canada, *Bull. Seism. Soc. Am.,* **82**, 2014-2031.

Atkinson, G. M., and D. M. Boore (1995). Ground-motion relations for eastern North America, *Bull. Seism. Soc. Am.,* **85**, 17-30.

Baptie, B., L. Ottemoller, S. Sargeant, G. Ford, and A. O'Mongain (2005). The Dudley earthquake of 2002: A moderate sized earthquake in the UK. *Tectonophysics,* **401**, 1-2: 1-22.

Bay, F., D. Fäh, L. Malagnini, and D. Giardini (2003). Spectral shear-wave ground-motion scaling in Switzerland, *Bull. Seismol. Soc. Am.*, **93**, 414-429.

Beresnev, I. A. and K.-L. Wen. (1996). Nonlinear soil response - a reality? *Bull. Seismol. Soc. Am.,* **86**, 1964-1978.

Boatwright, J. (1978). Detailed spectral analysis of two small New York State earthquakes, *Bull. Seismol. Soc. Am.*, **68**, 1117-1131.

Bommer, J. J., F. Scherbaum, H. Bungum, F. Cotton, F. Sabetta, and N. A. Abrahamson (2005). On the use of logic trees for ground-motion prediction equations in seismic

hazard analysis, *Bull. Seismol. Soc. Am.*, **95**, 377-389.

Bommer, J.J. and Crowley, H. (2006) The influence of ground motion variability in earthquake loss modelling. *Bulletin of Earthquake Engineering,* **4**(3), 249-273

Bommer, J. J. (2006). Empirical estimates of ground motion: advances and issues, *Proceedings of the Third International Symposium on the effects of Surface Geology on Seismic Motion, Grenoble, France*.

Bommer, J. J., P. J. Stafford, J. E. Alarcón, and S. Akkar (2007). Empirical ground-motion prediction over extended magnitude ranges. *Submitted to Bull. Seismol. Soc. Am.*

Boore, D. M. and J. Boatwright (1984). Average body-wave radiation coefficients, Bull. Seismol. Soc. Am. 74, 1615-1621.

Boore, D. M. (2003). Simulation of ground motion using the stochastic method, *Pure and Applied Geophysics,* **160**, 635-676.

Boore, D. M. (cited 06/2007). SMSIM - Fortran Programs for Simulating Ground Motions from Earthquakes, [online] "http://quake.usgs.gov/~boore/smsim/".

Brent, R. P. (1973). Algorithms for Minimization without Derivatives, *Englewood Cliffs, NJ: Prentice-Hall*, Chapter 7.

Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.*, **75**, 4997-5010.

Brune, J. N. (1971). Correction: Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.*, **76**, 5002.

Campillo, M., M. Bouchon, and B. Massinon (1984). Theoretical-study of the excitation, spectral characteristics, and geometrical attenuation of regional seismic phases, *Bull. Seismol. Soc. Am.*, **74**, 79-90.

Cormier, V. F. (1982). The effect of attenuation on seismic body waves, *Bull. Seismol. Soc. Am.*, **72**, 169-200.

Cotton, F., F. Scherbaum, J. J. Bommer, and H. Bungum (2006). Criteria for selecting and adjusting ground-motion models for specific target applications: applications to Central Europe and rock sites, Journal of Seismology, **10**, 137-156.

Cotton, F., Pousse G., F., Bonilla, L. and Scherbaum, F. (2008). On the discrepancy of recent European ground-motion observations and predictions from empirical models: analysis of KiK-net accelerometric data and point-sources stochastic simulations. Bull. Seism. Soc. Am, in press.

Cundall P.A. (1992). Theoretical basis of the program WAVE. Unpublished internal report, COMRO (now CSIR, Division of Mining Technology, South Africa), pp 1–12.

Deichmann, N. (2006). Local magnitude, a moment revisited, Bulletin of the Seismological Society of America **96**, 1267-1277.

Drouet, S., F. Cotton, and A. Souriau (2005). Attenuation, seismic moments, and site effects for weak motion events. Application to the Pyrenees, *Bull. Seismol. Soc. Am.*, **95**, 1731-1748.

Eberhart-Phillips, D. (1993). Local earthquake tomography: earthquake source regions, in Seismic Tomography: Theory and Practice, edited by Iyer, H.M., and Hirahara, K., pp. 613-643, Chapman and Hall, New York.

Edwards, B., A. Rietbrock, J. J. Bommer, and B. Baptie (2008). The Acquisition of Source, Path and Site Effects from Micro-earthquake Recordings using Q Tomography: Application to the UK, *submitted to Bull. Seismol. Soc. Am.*

Eshelby, J. D. (1957). The determination of an elastic field of an ellipsoidal inclusion and related problems, *Proc. Roy. Soc. London A*, **241**, 376-396.

European Seiesmological Commission (1998). European Macroseismic Scale (1998). EMS-98. Editor. G. Grünthal. Chairman of the ESC Working Group "Macroseismic Scales". GeoForschungsZentrum Potsdam.

Frankel, A. (1991). Mechanisms of seismic attenuation in the crust: scattering and anelasticity in New York State, South Africa, and southern California, *J. Geophys. Res.*, **96**, 6269 –6289.

Great Britain. Department of Trade and Industry. (2006). The Energy Challenge. CM6887. London: HMSO.

Great Britain. Department of Trade and Industry. (2007). Meeting The Energy Challenge: a white paper on energy. CM7124. London: HMSO.

Grünthal, G. and R. Wahlström (2004). An M_w based earthquake Catalogue for central, northern and northwestern Europe using a hierarchy of magnitude conversions, *Journal of Seismology*, **7**, 507-531.

Gutenberg, B. and C. F. Richter (1954). Seismicity of the Earth and Associated Phenomena, 2nd ed., Princeton University Press, Princeton, New Jersey.

Haberland, C., and A. Rietbrock (2001). Attenuation tomography in the western central Andes: a detailed insight into the structure of a magmatic arc, *J. Geophys. Res.*, **106**, 11151-11167.

Hanks, T., and H. Kanamori (1979). A moment magnitude scale. *J. Geophys. Res.*, **84**, 2348-2350.

Hanks, T. (1982). fmax. *Bull. Seismol. Soc. Am.* **72**, 1867-1879.

Hartzell S, L. F. Bonilla, and R. A. Williams (2004). Prediction of nonlinear soil effects, *Bull. Seismol. Soc. Am.*, **94**, 1609-1629.

Hashizume, M. (1979). Q of the crust beneath southwestern Honshu, Japan, derived from explosion seismic waves, *J. Phys. Earth. Planet. In.*, 20, 25-32.

Herrmann, R. B., and L. Malagnini (1996). Absolute ground motion scaling in the New Madrid Seismic Zone, *Seismol. Res. Lett.*, **67**, 40.

Hildyard M. W. (1995) WAVE developments. In: Develop a quantitative understanding of rock-mass behaviour near excavations in deep mines. Unpublished SIMRAC final project report(GAP029), December 1995. CSIR Mining Technology, Johannesburg, South Africa, pp 17–51.

Honda, R., S. Aoi, H. Sekiguchi, N. Morikawa, T. Kunugi, and H. Fujiwara (cited 02/2007). Ground Motion and Rupture Process of the 2004 Mid Niigata Prefecture Earthquake Obtained from Strong Motion Data of K-NET and Kik-net. [online] *"http://www.k-net.bosai.go.jp/k-net"*.

Horike, M. and T. Nishimura (2004), Attenuation Relationships of Peak Ground Velocity Inferred from the Kyoshin Network Data, *Journal of Struct. Construct. Eng.* (Transactions of AIJ), **575**, 73-79 (in Japanese).

Hutton, L. K., and D. Boore (1987). The M_L scale in Southern California. *Bull. Seism. Soc. Am.*, **77**, 2074-2094.

Jin, A., C. A. Moya, and M. Ando (2000). Simultaneous determination of site responses and source parameters of small earthquakes along the Atotsugawa fault zone, central Japan, *Bulletin of the Seismological Society of America*, **90**, 1430-1445.

Lees, J. M., and G. T. Lindley (1994). Three dimensional attenuation tomography at Loma Prieta: Inversion of t* for Q, *J. Geophys. Res.*, **99**, 6843-6863.

Lees, J. M. and J. Park (1995). Multiple-taper spectral analysis: a stand-alone C subroutine, *Computers and Geosciences*, **21**, 199-236.

Liu, H.P., D.L. Anderson, H. Kanamori (1976). Velocity dispersion due to anelasticity; implications for seismology and mantle composition, *Geophys. J. R. Astron. Soc.*, **47**, 41-58.

Madariaga R, K. Olsen, and R. Archuleta (1998). Modeling dynamic rupture in a 3D earthquake fault model, *Bull. Seismol. Soc. Am.*, **88**, 1182-1197.

Madariaga, R. (1976). The dynamics of an expanding circular fault, Bull. Seismol. Soc. Am., **66**, 639-666.

Malagnini, L., R. B. Herrmann, and K. Koch (2000). Regional ground-motion scaling in central Europe, *Bull. Seismol. Soc. Am.*, **90**, 1052-1061.

Marin, S., J. P. Avouac, N. Marc, and A. Schlupp (2004). A probabilistic approach to seismic hazard in Metropolitan France, *Bull. Seismol. Soc. Am.*, **94**, 2137-2163.

Murphy, J. R., Davis, A. H., and Weaver, N. L. (1971). Amplification of seismic body waves by low-velocity surface layers. *Bull. Seism. Soc. Am.*, **61**, 109-145.

Newmark, N. M. and W. J. Hall (1982). Earthquake Spectra and Design, Monograph prepared for Earthquake Engineering Research Institute, Berkeley, California.

Nishimura, T. and M. Horike (2003). The Attenuation Relationships of Peak Ground Accelerations for the Horizontal and the Vertical Components Inferred from Kyoshin Network Data, *Journal of Struct. Construct. Eng.* (Transactions of AIJ), **571**, 63-70 (in Japanese).

Ou, G.B., and R. B. Herrmann (1990). A statistical model for ground motion produced by earthquakes at local and regional distances, *Bull. Seismol. Soc. Am.*, **80**, 1397-1417.

Park, J., C. R. Lindberg, and F. L. Vernon (1987). Multitaper spectral analysis of high

frequency seismograms, *J. Geophys. Res.*, **92**, 12675-12684.

Pousse G., L. F. Bonilla, F. Cotton, and L. Margerin (2006). Nonstationary stochastic simulation of strong ground motion time histories including natural variability: application to the K-Net Japanese database, *Bull. Seismol. Soc. Am.*, **96**, 2103 - 2117.

Pozo, R. (2004). Template Numerical Toolkit. [online]. Available from: http://math.nist.gov/tnt/, [cited January 2007].

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1997). *Numerical Recipes in C: The Art of Scientific Computing Second Edition:* Cambridge University Press, pp. 278-281; pp. 288-290; pp. 417-424; and pp. 515-518 .

Principia Mechanica Ltd. (1982). *British earthquakes:* PML, Cambridge.

Principia Mechanica Ltd. (1988). *UK uniform risk spectra*: PML Report no HPC-IP-096013.

Raoof, M., R. B. Herrmann, and L. Malagnini (1999). Attenuation and excitation of three-component ground motion in southern California, *Bull. Seismol. Soc. Am.*, **89**, 888-902.

Rietbrock, A. (1996). Entwicklung eines Programmsystems zur konsistenten Auswer-tung grosser seismologischer Datensaetze mit Anwendung auf die Untersuchung der Absorptionsstruktur der Loma-Prieta-Region, Kalifornien. PhD-Thesis, University of Munich, pp. 141.

Rietbrock, A. (2001). P wave attenuation structure in the fault area of the 1995 Kobe earthquake, *J. Geophys. Res.*, **106**, 4141-4154.

Rietbrock, A., F. Scherbaum, F. Cotton, and D. Fäh (2004). On the determination of source, path, and site effects from microearthquake recordings for strong ground motion prediction. *Proceedings of the 29th ESC general assembly, Potsdam, Germany.*

Sabetta, F. and A. Pugliese (1996). Estimation of response spectra and simulation of nonstationary earthquake ground motions, *Bull. Seismol. Soc. Am.*, ***86,*** 337-352.

Salah, M. K., and D. Zhao (2003). Three dimensional attenuation structure beneath southwest Japan estimated from spectra of microearthquakes, *Physics of the Earth and Planetory Interiors*, **136**, 215-231.

Satoh, T. (2006). Inversion of Q_s of deep sediments from surface-to-borehole spectral ratios considering obliquely incident SH and SV waves, *Bull. Seismol. Soc. of Am.*, **96**, 943-956

Scherbaum, F. (1990). Combined inversion for the three-dimensional Q structure and source parameters using microearthquake spectra, *J. Geophys. Res.*, **95**, 12423-12438.

Scherbaum, F. (2001). Of Poles and Zeros: Fundamentals of Digital Seismology, second edition, *Kluwer Academic Publishers*, 265 pp.

Scherbaum, F., J. J. Bommer, H. Bungum, F. Cotton, and N. A. Abrahamson (2005). Composite ground-motion models and logic trees: methodology, sensitivities and uncertainties, *Bull. Seismol. Soc. Am.*, **95**, 1575-1593.

Steidl, J. H., A. G. Tumarkin, and R. J. Archuleta (1996). What is a reference site? *Bull.*

Seism. Soc. Am., **86,** 1733 –1748.

Suzuki, S. (1971). Anomalous attenuation of P waves in the Matsushiro earthquake swarm area. *J. Phys. Earth.*, **20**, 1-21.

Thurber, C. H. (1983). Earthquake locations and three-dimensional structure in the Cyote Lake area, central California, *J. Geophy. Res.*, **88**, 8226-8236.

Thurber, C. H. (1993). Local earthquake tomography: velocities and vp/vs theory, in Seismic Tomography: Theory and Practice, edited by Iyer, H. M., and Hirahara, K.: Chapman and Hall, New York, pp. 563-583.

Toro, G.R., N.A. Abrahamson and J.F. Schneider (1997). A Model of Strong Ground Motions from Earthquakes in Central and Eastern North America: Best Estimates and Uncertainties, *Seismological Research Letters*, **68**, 41-57.

Tsumura, N., S. Matsumoto, S. Horiuchi, and A. Hasegawa (2000). Three-dimensional attenuation structure beneath the northeastern Japan arc estimated from spectra of small earthquakes, Tectonophysics, **319**, 241–260.

Utsu, T. (2002). "A list of deadly earthquakes in the World: 1500-2000". *International handbook of earthquake engineering and seismology, Academic Press* 81A: 691-717.

Wassermann, J. (2002). IASPEI New manual of seismological observatory practice, vol. 1, edited by P. Bormann, *GeoForschungsZentrum Potsdam*, 42 pp.

Wessel, P. and W. H. F. Smith (1991). Free software helps map and display data. EOS Trans. A*m. Geophys. Union,* **72**, pp. 441 and pp. 445–446.

Wessel, P., and W. H. F. Smith (1998). New, improved version of the Generic Mapping Tools Released, *EOS Trans. AGU, 79*, 579.

Index of Figures **Page**

Page

Page

Index of Tables

Page

Appendices

CSPECMC Manual

NAME

cspecmc - Converts a database of time-series seismic records into a database of Fourier spectra for use with SFREQFIT and SAMPFIT. Includes options for windowing, tapering and processing of the time series and can process GSE, SEISAN, KIK-NET, ISSED and SMC formats.

SYNOPSIS

cspecmc database.dat -W<type/ndirs/./path/> -H<type/./path/> [options]

[...

-a -w<noise_len/sig_len> -c<taper> -p<p_id/search> -t<win_offset> -q -g<velo/win> -G<dist2/vel2/dist1/vel1> -m<f_type> -d<d_type> -x<buffer/./path/> -M<time/buffer> -P -I<min/max> -o -u<./path/> -C<d_type> -U -X -r -E<min>

...]

AVAILABILITY

All UNIX flavours

DESCRIPTION

cspecmc processes databases of recorded time-series and phase information to provide spectral data in the format of SFREQFIT. Seismic data can be stored in a variety of formats. CSPECMC can process GSE, SEISAN, KIK-NET, ISSED and SMC data stored in any directory structure. A variety of windowing options are provided, capitalising on phase information (NORDIC format or GIANT format), or peak amplitudes. Input data can be velocity, accleration or displacement time series. CSPECMC can output velocity, accleration or displacement Fourier spectra (this does not have to be the same as the input type). The windowed (or whole) time-series data used in producing the Fourier spectra may also be output in units of velocity, accleration or displacement (again, independant of the input or output of Fourier spectra). CSPECMC can correct for instrument response (using SEISAN or GSE response data) providing true ground motion data.

REQUIREMENTS

database.dat

This can be either: a NORDIC format index of data locations, event information and phase data; or a list of time-series data locations.

-W<type/ndirs/./path/>

this informs the programme the data 'type' (GSE, SEISAN, Kik-Net etc.) and the 'path' to the root of the data. 'ndirs' indicates the number of directories that precede the filename specified in 'database.dat' so that the filename can be separated from the string given in 'database.dat'.

-H<type/./path/>

OPTIONS

Usage: cspecmc <select.out>

Options:

-w<noise/signal>

Specify noise and signal window lengths

-a Print all spectra to file all_spectra.out

-c<20%>
Specify cosine taper design

-p<1|2|3|4>
Specify Phase for signal specra:
1: Phase P
2: Phase S
3<a>: No phase information. Append 'a' and use -o
to search for max ampl. Otherwise signal time is set to event time.
4: Giant format picks used.

-t<offset>
Specify Signal window offset in seconds

-d<d|v|a>
Specify output data type:
d: Displacement (m) Spectra
v: Velocity (m Spectra
a: Acceleration (m Spectra

-m<1>
Use multitaper routines

-M<time/buffer>
Extend window to include maximum amplitude up to 'time' sec and add 'buffer' sec after max.

-g<vel/time>
Use group velocity (vel) to define Signal window end
define window start with time => win length before Signal
arrival at 50KM (increases linearly with dist. Defaults 2.8/1

-G<dist/vel>
Use differential group velocity (vel) to define
Signal window end beyond (dist)

-x<time/path>
print frequency and time (S_time - 5 sec to S_time + (time) sec
data series to path.

-C<d|v|a>
Correct waveform time series using instrument response.

-u<./path/>
Full path to station response file directory.
nb. if file not available or no suitable calibration file found record is removed

-o Center (+/- offset) signal window on maximum (uncorrected) amplitude.

-H<htype/dir>
Specify alternative input formats:
htype:
1 = Nordic S-file formatted index;
2 = Giant directory structure format (see man page).

-W<wtype/dir>
Specify alternative input formats: wtype:
1 = Nordic format (binary) waveform files;
2 = GSE format (binary) waveform files;
3 = KiKNET format (ascii);
4 = PDAS format (binary) waveform files.
5 = MiniSEED format (binary) waveform files.

6 = ISESD ascii format waveform files.
7 = SCM ascii format waveform files.

-r Don't check response file continuity (for naming differences in files).

-U Use gain at natural period of seismometer (e.g., if no CAL files)
nb. If not available record will be set to 0!

-? This help.

INPUT FILE FORMATS

database.dat

This file must be the first argument after the CSPECMC command, otherwise, or if the file cannot be found, the programme will exit. If not a NORDIC format file (SEE ALSO) the file should list the data files, each separated by a new-line character. For example, if the data files are in a database format of ./yyyy/mm/data.ext , use the command: *ls ./????/??/*.ext > database.dat* will probably be sufficient to create the correct format as long as there are no unwanted files in the ./????/??/ directory structure. The 'ndirs' must be specified on the -W flag. In this case it is 2 (e.g., ???? and ?? directories).

Example contents:

1997/07/data1.sac
1997/08/data2.sac
1997/11/data3.sac
1998/02/data4.sac
2001/05/data5.sac

The programme processes each file in this list (or each file found in the NORDIC file format) in order. Each time-series file should, unless in SEISAN format, contain only one time-series (e.g., a three component instrument will need three files for the separate components). In SEISAN format the data files may contain any number of instruments and components from a single event, or part there of.

The formats of acceptable input data may be found by referring to the SEE ALSO section of this manual.

stn_cal_index.dat

Similarly to the format specified for the "database.dat" file, the "stn_cal_index.dat" file lists the instrument response files separated by a newline. The path (from the location given in 'path' on option -u) must be given. E.g., if the file is located at "/ABCD/001/file.dat" and the option "-u/ABCD/" is given: the location in "stn_cal_index.dat" must be "001/file.dat".

STATION.HYP

This is a list of station location data in the HYPO71 format beginning on the TWELTH line of the file (the first 11 lines are ignored by the programme):

c-type format: " %4s%2d%5.2f%c%3d%5.2f%c%4d"
E.g.,

line 12 " LRW 6008.16N 110.67W 98
line 13 SAN 6001.07N 114.35W 150
line 14 WAL 6015.38N 137.04W 167
line 15 YEL 6033.05N 104.98W 203

```
.                .
.                .
.                .
line 100  ABC 6035.55N  103.48W  33
line 101
line 102  Other information..."
```

The final line in this file must be a character return (blank line) or simply the end of the file: Any data at the end of the file not in this format (comments etc) must be separated by a blank line.

The file is not required but it is useful if station data are not included in the headers of the time-series data: if included the program will search the "STATION0.HYP" file and extract location data from it. This data will take precedence over any other location data in the headers of the time-series. Hence if the latter is more reliable DO NOT include this file in the working directory.

The actual HYPO71 format may include some more information in the first 12 lines of the file and after (to the right) of the last digits in the example shown above, these can be left in the file but will not be used.

Time-Series data must be in GSE, SEISAN, KIK-NET, ISSED or SMC format (see SEE ALSO).

Instrument Response Data must be in SEISAN Response or GSE response format (see SEE ALSO).

Key:

#	event header type indicator
yyyy	4 digit integer year
mm	2 digit integer month (2 digit numbers precede 0 if too short: e.g., 01, 02, etc.)
dd	2 digit integer day
hh	2 digit integer hour
mi	2 digit integer minute
ss.ss	floating point seconds
deg_lat	floating point degrees degrees latitude
min_lat	floating point minutes degrees latitude
deg_lon	floating point degrees degrees longitude
min_lat	floating point minutes degrees longitude
depth	focal depth (km)
magnitude	magnitude of event
mag_type	magnitude measurement type: e.g., L = ML type
stat_code	4 character station code
comp_code	4 character component code
blank	zero
phase_ttime	Travel time from event to recording (s)
azimuth	source-station azimuth (degrees)
hypo_dist	hypocentral distance (km)

sig_mom	observation signal moment (m.s)
t-star	obervation t* value (s)
corner_freq	source corner frequency (Hz)
alpha	frequency dependance of Q given by Q(f)=Q0.f^alpha for record.
freq_start	first frequency (Hz)
df	frequency interval (Hz)
win_length	window length used (sec)

Output Data Files

Each event has its own file. Any records of this event are stored in order in this file. Each observation has two header lines: an event header (column one is "#") and an observation header (column one is "@"). Following the two header lines is the amplitude/velocity/acceleration spectrum in two columns: signal amplitude and noise amplitude.

Data files are not particularly dependant on formatting (the program distinguishes data using whitespace - gaps between numbers), but the programme requires all data to be present: if missing data it should be input as a value such as -99999 or similar. A notable exception is the header indicators @ and # which must be placed in column 1, and the date/time format which is described below.

Generic data file:

yyyymmdd hhmi ss.ss deg_lat min_lat deg_lon min_lon depth corner_freq magnitude mag_type
@ stat_code comp_code azimuth phase_ttime hypo_dist sig_mom corner_freq t-star alpha freq_start df win_length
signal_ampl noise_ampl
signal_ampl noise_ampl
signal_ampl noise_ampl
.
.
.
freq signal_ampl noise_ampl
yyyymmdd hhmi ss.ss deg_lat min_lat deg_lon min_lon depth corner_freq magnitude mag_type
@ stat_code comp_code azimuth phase_ttime hypo_dist sig_mom corner_freq t-star alpha freq_start df win_length
signal_ampl noise_ampl
signal_ampl noise_ampl
.
.
.

Example data file:

20050120 2213 23.000000 56.490002 0.000000 -4.379000 0.000000 5.400000 0.0 2.700000 L
@ GAL_ S__Z 0.0 48.510002 182.082260 0.0 0.0 0.0 0.0 0.048843 0.048843 4.0
8.43677e-06 1.77636e-05
2.23235e-06 4.59214e-06
1.01664e-06 2.14286e-06
6.16222e-07 1.26403e-06

COMMON EXAMPLE

cspecmc index_UD2.dat -w4/10 -t2 -E6 -p3a -dv -m1 -W3#2#../data/ -H2 -U -o -P -I5/95 -x5#./vdata+-10/

this command searches the file "index_UD2.dat" for time-series data (-H2) in Kik-Net format (-W3) which reside in a 2 deep directory structure (-W3#2) at ../data/ (-w3#3#../data/). A signal window of 10 seconds and noise of 4 seconds (-w4/10) are created. There is no phase information (-p3), or reliable event time information so we start the signal time so that the window will fit in the trace (-p3a). The window is then shifted to be centred 2 second from the maximum (-o -t2). If the window does not fit inside the trace (starts before or ends after the start or end of the trace respectively) we allow the window to be reduced to a minimum of 6 seconds (-E6). The signal window is then reduced to remove the fist and last 5% of the energy in the window (-I5/95). A multitaper (-m1) Fourier transform is applied and the spectrum is converted to velocity (-dv). The original windowed data (in counts as no -C option was used) is output to the directory ./vdata/ with 5 second frame around the window. The -U tells the programme to apply the gain found in the headers to the data. If the gain is not included or is 0 this will render the data useless and it will not be included.

SEE ALSO

sfreqfit(1), sampfit(2),

Edwards, B., A. Rietbrock, J. J. Bommer, and B. Baptie (2007). The Acquisition of Source, Path and Site Effects from Micro-earthquake Recordings using Q Tomography: Application to the UK, submitted to Bull. Seismol. Soc. Am.

SMC National Strong-Motion Program; 2002: SMC-format data files. On Internet at: http: //nsmp.wr.usgs.gov/smcfmt.html.

KIK-NET Kyoshin Net; 2002: About K-NET data format. On Internet at: http: //knetwww.k-net.bosai.go.jp/k-net/man/knetform_en.html.
http://www.k-net.bosai.go.jp/k-net/man/knetform_en.html
http://www.kik.bosai.go.jp/kik/man/knetform_en.html

SAC Goldstein P.; 2002: SAC data file format. On Internet at: http: //www.llnl.gov/sac/.

ISESD Ambraseys et al. (2004): isesd.cv.ic.ac.uk/ESD/ambraseysetal2004bgta.pdf

GSE www.seismo.ethz.ch/autodrm/downloads/provisional_GSE2.1.pdf

SEISAN Jens Havskov and Lars OttemÃ¶ller: www.geoinstr.com/pub/manuals/seisan_7.2.pdf

HYPO71 (for STATION0.HYP format) http://jclahr.com/science/software/hypo71/hypo71manual.pdf

BUGS

Please email author.

AUTHOR

Benjamin Edwards - b.edwards@liv.ac.uk

SFREQFIT Manual

NAME

sfreqfit - finds best fitting spectral models to observed or synthetic seismic spectra.

SYNOPSIS

sfreqfit <event_index.dat> -[r|s]<data_type> -x<inversion_type> [options]

[...

-m<min/max> -c<min/max> -n<snr> -N -R -d<min/max> -A<{A/D}> -y<{S/A/B/H}> -z<{Z/N/E/H{1|2|3}[i]}> -o<freq_band[a]> -Z<dist/alpha/Q> -S<falloff> -f<num> -F<start{/|s}end{/|s}inc> -w<fc window> -a<alpha> -I</data path/> -J</data path/> -e</data_file> -i</data path/> -t<Q> -T</t* data file> -p -q</data path/> -E<velo> -B<freq/ropt> -b<nsd/cut> -g<mag_min/mag_max> -C<fc_int/fc_grad/{file|0/mw_int/mw_grad1/mw_grad2/mw_grad3}> -l -D -? -h

...]

The character '/' is used as a spacer and can usually be repaced with any non-numeric character. Sometimes the spacer is used to imply how the option is constructed or to provide more details e.g., for -F<start{/|s}end{/|s}inc>; if using '/' the programme assumes you entered corner frequency, if 's' is used between the numbers it assumes you are using stress drop. This will be explicitly stated in OPTIONS if it is the case. See COMMON EXAMPLES for examples of useage if you are unsure.

All variables within < > are required. Using an option without entering all relevant values will lead to unpredictable behaviour. One or more variable within { }brackets is required. Variables within [] are optional. 'l' is an OR operator: choose only one.

AVAILABILITY

All UNIX flavours

DESCRIPTION

SFREQFIT finds best fitting spectral models to observed or synthetic seismic spectra based on data processing and inversion options used at the command line. An index file e.g., "event_index.dat" must list all the data files to be processed. Each data file must contain all of the records from one event. This structure allows large databases to be quickly processed. Detailed input and output file formats can be found at the end of this document (under FILE FORMATS). A variety of forward and inverse modelling techniques are available to use. The programme also alows the use of results from SAMPFIT in order to specify certain parameters.

REQUIREMENTS

-r or -s

use -r<d|v|a> for real data, -s<d|v|a> for synthetic (or bootstrap) data outputs. select data format: v (velocity), a (acceleration) or d (displacement).

-x select inversion type: 1 (L1 absolute amplitude), 2 (Fast iterative L2 (least squares)), 3 (Maximum Likelihood) or 4 (non-linear L2), 5 (log-space L2) or 6 (log-space L1).

OPTIONS

words within apostrophes correspond to those used in the SYNOPSIS (to refer to for precise usage).

Data Selection and Processing Options

-m specify the range between 'min' and 'max' frequencies to pass to data processing stage. This is useful to avoid frequencies above or below filters. E.g., -m1/20

-c specify 'min' and 'max' frequencies (Hz) between which the SNR (signal-to-noise ratio) restriction must be met. If the SNR is not met at low frequencies the minimum used frequency is increased by one step to a maximum of 'min'; the same applies to the maximum frequency used, with the maximum frequency decreasing to a minimum of 'max' Hz. E.g., -c2/10

-n specify the signal-to-noise ratio ('snr') criteria for accepting data. This must be used with the -c option to define a frequency band over which the SNR must be met. E.g., -n3

-d filter observations to include only those between 'min' and 'max' km in hypocentral distance. E.g., -d10/100

-y filter observations to include only those from: S (velocimeters), A (accelerometers), B (broadband 80Hz), H(broadband 100Hz) instruments. E.g., -yS/B/H

-A specify whether to include A (analogue) and/or D (digital) instruments. E.g., -AD

-z filter observations to include only those from: Z (vertical), N (north), E (east) or H (both north and east to form horizontal geometrcal mean). E.g., -zN/E or -zZ etc. Append H with 1, 2, or 3 to specify geometrical mean, vector sum or largest of both horizontal components. Append H{1|2|3} with 'i' e.g., -H2i to ignore the exact orientation given in the data; any two consecutive data from the same earthquake to the same station are considered a horizontal pair: useful if the orientation was given in degrees and the conversion in CSPECMC failed to assign an exact N/S/E/W label but the format is always consecutive N-S/E-W data.

-o smooth input spectra over frequency bands of 'freq_band' Hz. Append with 'a' (e.g., -o4a) to use an automatic shortening of this band at the ends of the spectrum.

-E estmiate travel times using hypocentral distance and an average velocity 'velo' in m/s: required if the programme CSPECMC did not give the exact travel time. E.g., -E3500

-B enable bootstrap mode. Each record has a 'freq'% chance of being used. If 'ropt' is 'R' (e.g., -B80R) then the unused record is dicarded. If not (-B80) the unused record is replaced by a random record from the same event.

-b check signal moments for consistency after processing each event: if a signal moment, normalised by 1/R to its source is greater than 'nsd' standard deviations from the mean the record is removed. The standard deviation and mean are recalculated and the process repeats until the fractional change in the standard deviation is less than 'cut'. This can help to remove bad data from an event.

-g discard records from earthquakes with magnitudes above 'mag_max' or below 'mag_min'. E.g., -g3/7

-N remove the noise estimate from the record using (S(f)^2-N(f)^2)^0.5 where S(f) is the signal spectrum and N(f) is the noise spectrum.

-R remove unresolved records, defined by the optimum model being on the extrema of the grid-search.

Inversion options

-t fix Quality factor at 'Q' for all records (homogeneous Q). E.g., -t400

-T use simul2000 output file (e.g., f27) or SHOT file to search for t* values for individual records: use for 1D, 2D or 3D Q models. E.g.: Use for 3 stage spectral fit: (1)SFREQFIT -> (2)SIMUL2000 -> (3)SFREQFIT

1. Run SFREQFIT first to construct a "tstar.simul" file (known as "SHOT" file in simul2000) without the -t or -T options;
2. rename the new "tstar.simul" file 'SHOT', setup the remaining simul2000 files ("CNTL", "STNS", etc.: see manual), then invert for a Q model by running simul2000;
3. take your new Q model and use simul2000 in forward modelling mode to produce an f27 output

file which contains t* computations for all records you initially produced in tstar.simul. This f27 file can now be used to fix t* in SFREQFIT. E.g., -T./dir/f27

-Z treat surface waves separately. Specify distance 'dist' where surface waves exist beyond, frequency dependance 'alpha' and Q values (if applicable) where Q(f)='Q'.f**'alpha'. Q values are only fixed if -t or -T options are also used. E.g., -Z150/0.0/400

-S specify source spectrum shape: -1 (boatwright), 2 (brune omega squared), 3 (omega cubed) or any other number x.x for an omega x.x spectrum. E.g., -S2

-f force a common corner frequency for each event with more than 'num' records. Discard events with less than 'num' records.

-F specify the grid search resolution with 'min', 'max' and 'inc'rements of the source corner frequency. If the separator of the min and max is 's' e.g., -F1s100/0.5 the 'min' and 'max' values are input in bars of stress drop. If the increment is appended % e.g., -F1s100/10% the increment is 10% of the current corner frequency.

-w specify the width (in Hz), about the spectral peak, for which to search for the corner frequency. Corner frequencies outside this range are skipped. E.g., -w10

-a specify the frequency dependance of Q as Q=Q.f**'alpha'. If this option is not included at runtime the programme will invert for alpha on a spectrum-by-spectrum basis (i.e. as a free parameter between 0 and 1). E.g., -a0.1

-C forward model the corner frequency using the relation log(fc)= 'fc_int' + 'fc_grad' * Mw, where the moment magnitude of a record is found in file 'file' OR the magnitude is computed using the polynomial Mw = 'mw_int' + 'mw_grad1'*ml + 'mw_grad2'*ml^2 + 'mw_grad3'*ml^3 if the number '0' is used instead of a filename (see COMMON EXAMPLES). The file must be in the output format of SAMPFIT (see its manual entry for details). E.g., -C2/-0.5/0/0/1/0/0 or -C2/-0.5/./dir/dir/smom_corr.out

-M fix signal moments using the model output from the SAMPFIT output file. E.g., -M./dir/dir/smom_corr.out

-e do not perform inversion. Use 'data_file' (results of amplitude inversion SAMPFIT "smom_corr.out") to find final misfits from all models (frequency and amplitude). E.g., -e./dir/dir/smom_corr.out

Output options

-I output noisy synthetic data based on inversion results with addition of original noise spectrum into the directory '/data_path/'. E.g., -I./dir/dir/

-J output original data, with inversion results included as synthetic data to '/data_path/'. Use for e.g., bootstrap analysis. E.g., -J./dir/dir/

-i output fractional residuals data/model to individual files for constructing site reponses. E.g., -i./dir/dir/

-p output grid-search inversion results to "inversion_iter.out".

-q output model and data for each record to '/data_path/'. Useful for plotting data against model results. E.g., -q./dir/dir/

-l compute RMS (root-mean-square) misfit of the model for each spectrum.

-D enable debug mode: progress bar is replaced with stderr output (which usually redirects to output.dat).

INPUT FILE FORMATS

Key:

#	event header type indicator
yyyy	4 digit integer year
mm	2 digit integer month (2 digit numbers precede 0 if too short: e.g., 01, 02, etc.)
dd	2 digit integer day
hh	2 digit integer hour
mi	2 digit integer minute
ss.ss	floating point seconds
deg_lat	floating point degrees latitude
min_lat	floating point minute part latitude
deg_lon	floating point degrees longitude
min_lat	floating point minute part longitude
depth	focal depth
corner_freq	source corner frequency of event
magnitude	magnitude of event
mag_type	magnitude measurement type: E.g., L or 1 for ML type
@	observation type header indicator
stat_code	4 character station code
comp_code	component code
azimuth	azimuth (degrees)
phase_ttime	travel time of signal (sec)
hypo_dist	hypocentral distance (km)
sig_mom	signal moment of observation (m.s)
corner_freq	source corner frequency of observation (Hz)
t-star	t* parameter of observation (sec)
freq_start	first frequency (Hz)
df	frequency interval (Hz)
win_length	window length used (sec)
sig_ampl	Fourier amplitude/velocity/acceleration (dependent on options) of signal
noise_ampl	Fourier amplitude/velocity/acceleration of noise
event_fit	average fit value (dependent on inversion type selected) for event
event_rms_	average rms fit value for event
Q	quality factor
f_start	minimum frequency (Hz) of bandwidth used in inversion
f_end	maximum frequency (Hz) of bandwidth used in inversion
obs_fit	fit value for observation (dependent on inversion type)
obs_rms	rms fit value for observation (dependent on inversion type)

Index File

This file must be the first argument after the SFREQFIT command, otherwise, or if the file cannot be found, the programme will exit. The file should list the data files, each separated by a new-line character. For example, if the data files are in a database format of ./yyyy/mm/yyyymmddhhmmss.ss, use the command:
ls ./????/??/ > event_index.dat*
to create the correct format as long as there are no other files in a ./????/??/ directory structure.

Example contents:

./1997/07/19970722000730.19
./1997/08/19970822000843.25
./1997/11/19971122001521.64
./1998/02/19980222001521.64
./2001/05/20010522001958.36

Data Files
Each event has its own file. Any records of this event are stored in order in this file. Each observation has two header lines: an event header (column one is "#") and an observation header (column one is "@"). Following the two header lines is the amplitude/velocity/acceleration spectrum in two columns: signal amplitude and noise amplitude.

Input data files are not particularly dependent on formatting (the program distinguishes data using whitespace - gaps between numbers), but the programme requires all data to be present: if missing data it should be input as a value such as -99999 or similar. A notable exception is the header indicators @ and # which must be placed in column 1, and the date/time format which is described below.

No degree of precision is implied by the numbers (e.g., 5.0000001) in the results files: the user should be aware of this and round-off accordingly.

Generic data file:

yyyymmdd hhmi ss.ss deg_lat min_lat deg_lon min_lon depth corner_freq magnitude mag_type
@ stat_code comp_code azimuth phase_ttime hypo_dist sig_mom corner_freq t-star alpha freq_start df win_length
signal_ampl noise_ampl
signal_ampl noise_ampl
signal_ampl noise_ampl
.
.
.
freq signal_ampl noise_ampl
yyyymmdd hhmi ss.ss deg_lat min_lat deg_lon min_lon depth corner_freq magnitude mag_type
@ stat_code comp_code azimuth phase_ttime hypo_dist sig_mom corner_freq t-star alpha freq_start df win_length
signal_ampl noise_ampl
signal_ampl noise_ampl
.
.
.

Example data file:
20050120 2213 23.000000 56.490002 0.000000 -4.379000 0.000000 5.400000 0.0 2.700000 L

@ GAL_ S__Z 0.0 48.510002 182.082260 0.0 0.0 0.0 0.0 0.048843 0.048843 4.0
8.43677e-06 1.77636e-05
2.23235e-06 4.59214e-06
1.01664e-06 2.14286e-06
6.16222e-07 1.26403e-06

OUTPUT FILE FORMATS

Inversion results file results.out

Generic Format:

yyyymmdd hhmi ss.ss deg_lat min_lat deg_lon min_lon depth event_fit event_rms magnitude mag_type
stat_code comp_code hypo_dist sig_mom t-star alpha corner_freq Q f_start f_end obs_fit obs_rms azimuth phase_ttime
stat_code comp_code hypo_dist sig_mom t-star alpha corner_freq Q f_start f_end obs_fit obs_rms azimuth phase_ttime
stat_code comp_code hypo_dist sig_mom t-star alpha corner_freq Q f_start f_end obs_fit obs_rms azimuth phase_ttime
yyyymmdd hhmi ss.ss deg_lat min_lat deg_lon min_lon depth event_fit event_rms magnitude mag_type
stat_code comp_code hypo_dist sig_mom t-star alpha corner_freq Q f_start f_end obs_fit obs_rms azimuth phase_ttime
.
.
.

Example results.out file:

19920227 0250 24.90 55.21 0.00 -3.42 0.00 5.90 2.93068e-13 5.05448e-07 2.70 L
XAL_ S_GZ 86.54 5.09993e-08 0.0207 0.000 22.491179 1215.942017 1.07 24.91 2.5988e-13 5.09784e-07 25.17 116.91
XDE_ S_GZ 79.07 7.34799e-08 0.0432 0.000 22.491179 540.277771 1.07 24.91 5.11702e-14 2.26208e-07 23.34 183.33
XSO_ S_GZ 80.57 8.87064e-08 0.0135 0.000 22.491179 1706.666748 1.07 24.91 7.29579e-13 8.54154e-07 23.04 67.21
19920228 0050 27.40 53.33 0.00 -1.18 0.00 11.60 2.46551e-13 4.96539e-07 2.00 L
LHO_ S_GZ 52.42 2.76217e-08 0.0174 0.000 34.891197 867.816101 1.47 24.91 2.46551e-13 4.96539e-07 15.10 298.49
.
.
.

Moment.dat output file (SAMPFIT input file)

Generic Format:

yyyymmdd hhmi ss.ss deg_lat min_lat deg_lon min_lon depth corner_freq magnitude mag_type
stat_code comp phase_ttime azimuth hypo_dist sig_mom t-star alpha corner_freq
stat_code comp phase_ttime azimuth hypo_dist sig_mom t-star alpha corner_freq
stat_code comp phase_ttime azimuth hypo_dist sig_mom t-star alpha corner_freq
.
.
.
stat_code comp phase_ttime azimuth hypo_dist sig_mom t-star corner_freq
yyyymmdd hhmi ss.ss deg_lat min_lat deg_lon min_lon depth corner_freq magnitude mag_type

stat_code comp phase_ttime azimuth hypo_dist sig_mom t-star alpha corner_freq
.
.
.

Example format:

```
# 19920227 0250 24.90 55.21 0.00 -3.42 0.00 5.90 22.49 2.70 L
XAL_ S_GZ 25.17 116.91 86.54 5.09993e-08 0.020700 0.00 22.49
XDE_ S_GZ 23.34 183.33 79.07 7.34799e-08 0.043200 0.00 22.49
XSO_ S_GZ 23.04 67.21 80.57 8.87064e-08 0.013500 0.00 22.49
# 19920228 0050 27.40 53.33 0.00 -1.18 0.00 11.60 34.89 2.00 L
LHO_ S_GZ 15.10 298.49 52.42 2.76217e-08 0.017400 0.00 34.89
.
.
.
```

tstar.simul Output File (simul2000 input)

A SHOT file is created by SFREQFIT in the input format of simul2000. See simul2000's manual for more details.

COMMON EXAMPLES

1. sfreqfit ./file_list.out -d10/300 -f3 -F0.01s1000/10% -a0 -rv -c2/20 -m1/25 -n3 -x5 -S2
Use real velocity data (-rv) recorded between 0 and 300km (-d10/300) if there are 3 or more recordings (-f3) of the earthquake which have a signal to noise ratio of 3 or more (-n3) between at least 2 and 20Hz (-c2/20). Only consider data between 1 and 25Hz (-m1/25): discard other frequencies. Fix the frequency dependance of Q at 0.0 (-a0.0). if -a? is left out the programme will invert for this alpha parameter for each record. We are using inversion type 5 (-x5) which is the lof-space L2 inversion and a Brune type source (-S2). The grid-search through the corner frequency is between the corresponding stress drop of 0.01bar and 1000bars with a 10% change in corner frequency per itereation (-F0.01s1000/10%).

SEE ALSO

sampfit(1), cspecmc(1),

Edwards, B., A. Rietbrock, J. J. Bommer, and B. Baptie (2008). The Acquisition of Source, Path and Site Effects from Micro-earthquake Recordings using Q Tomography: Application to the UK, submitted to Bull. Seismol. Soc. Am.

BUGS

Please email author.

AUTHOR

Benjamin Edwards - b.edwards@liv.ac.uk

SAMPFIT Manual

NAME

sampfit - Inverts for seismic moments, site amplification and geometrical spreading models from recorded signal moments.

SYNOPSIS

sampfit <moment.dat> <dist1>/<dist2>/<dist3> <cutoff> [options]

[...

-s<surf_dist> -f<decay1/decay2/decay3/decay4> -g<surf_decay> -o<stress_drop> -O<mw_int/mw_grad> -n<min_obs/min_stn> -r<source_size> -R -e -l -m -w -M<min_mag/max_mag> -S<min_mag/max_mag> -D<min_dist/max_dist> -A<azi_var> -c<c_factor> -I<max_iter> -v -F<decay> -V<velo_model> -z -X<ori> -h

...]

AVAILABILITY

All UNIX flavours

DESCRIPTION

Computes seismic moments, site amplification and a geometrical spreading model from signal moments. An input file is provided by the programme SFREQFIT (default:moment.dat). Runtime requirements and options are listed below. References to names in 'apostrophes' refer to the SYNOPSIS.

REQUIREMENTS

<moment.dat>

input file in the format provided by the programme SFREQFIT (see INPUT FILES section).

<dist1>/<dist2>/<dist3>

Set geometrical spreading model segmentation distances (km) - a maximum of 4 segments. Set higher than the furthest record if not required (e.g., 10/100/100000 for 3 segments, 10/100000/100000 for two segments).

<cutoff>

Set variance cutoff level: Observations falling outside 'cutoff' standard deviations from the mean are removed. Set high to disable.

OPTIONS

-s specify distance 'surf_dist' beyond which observations are classed as surface waves. These records are treated independantly in the geometrical spreading model with a separate 1/R**X section.

-f specify geometrical spreading model ('decay?' is the exponent of decay for the ?th section of the model) with up to four sections. The geometrical spreading model is not inverted for in this case.

-g as -f but specify single section model specifically for surface waves. Must also specify -s.

-o specify 'stress_drop'. Moments are forward computed assuming a circular rupture and the source corner frequency in the input file.

-O specify a linear equation for computing moments based on mw = 'mw_int' + 'mw_grad' * source corner frequency.

-n specify the minimum number of records of each earthquake 'min_obs' for the earthquake to be included in the inversion and the minimum number of records recorded a station for the station to be included in the inversion. Set 'min_stn' to 0 to switch off inversion for site amplification (set amplification factor = 1) while still including the data from those stations.

-r specify the source diameter 'source_size' if not equal to 1km. If this option is used some options are disabled due to their dependance on the 1km source size. This option is not really necessary as we convert to a point source when computing the moments.

-R compute the source size from the source corner frequency. Again this disables some useful features and provides no more functionality as we later convert to a point source. However, it is included for the possibility of further expansion.

-F fix the first segment of the geometrical decay model as 1/R**decay.

-e define that the last segment of the geometrical spreading model (segment 4) is linked to segment 1: they share the same decay value.

-l link the surface geometrical spreading model to the body geometrical spreading model at 'surf_dist': so there is overlap between the models. A body wave at 'surf_dist' will therefore be attenuated to the same degree as a surface wave at 'surf_dist'.

-m minimise (least-squares) the percentage difference between the data and model in the log-space (default is to minimise (least-squares) the absolute values in the log-space).

-w weight the signal moments by the value of the signal moment: records of closer or stronger earthquakes are assumed more reliable.

-v weight signal moments by their record distance. Records further away are assumed more reliable.

-M define minimum 'min_mag' and maximum 'max_mag' earthquake magnitudes to include for body waves.

-S as for -M but for surface waves.

-D specify minimum 'min_dist' and maximum 'max_dist' record distances to include.

-A define the standard deviation of azimuths for an event to be included. e.g., -A80 requires an azimuth standard deviation of 80 degrees for events to be included.

-c apply a constant correction factor to signal moments. E.g., if signal moments are in cgs use 0.01 to convert to SI.

-I define the maximum number of iterations before exiting.

-V use a velocity model in file 'velo_model' to estimate rupture velocity. The velocity model is simply depth and corresponding velocity separated by return values:

-z automatically load all available stations in the input file. If selecting stations an array of acceptable station names can be used to filter the input file. This array must be added in the source code however.

-X select orientation of the data you are using: 'ori' should be H or V for horizontal or vertical respectively.

INPUT FILE FORMATS

Key:

#	event header type indicator
yyyy	4 digit integer year
mm	2 digit integer month (2 digit numbers precede 0 if too short: e.g., 01, 02, etc.)
dd	2 digit integer day
hh	2 digit integer hour
mi	2 digit integer minute
ss.ss	floating point seconds
deg_lat	floating point degrees latitude

min_lat	floating point minute part latitude
deg_lon	floating point degrees longitude
min_lat	floating point minute part longitude
depth	focal depth
magnitude	magnitude of event
mag_type	magnitude measurement type: e.g., L = ML type
stat_code	4 character station code
blank	zero
azimuth	azimuth
hypo_dist	hypocentral distance
sig_mom	observation signal moment
t-star	obervation t* value
record_mag	magnitude recorded with input data
computed_mag	magnitude computed
phase_ttime	travel time used in computations
real_smom	recorded signal moment
model_smom	recomputed signal moment
fr_diff	%/100 difference between data and model
M3norm_smom	real signal moment normalised to a magnitude 3 event with no station amplication (for ploting geometrical spreading).
corner_freq	source corner frequency
alpha	frequency dependance of Q given by Q(f)=Q0.f**alpha for record.

Moment.dat Input Data

This file is the input data for the programme. It is the output of the programme SFREQFIT which produces signal moments from a database of Fourier spectra.

Generic Format:

```
# yyyymmdd hhmi ss.ss deg_lat min_lat deg_lon min_lon depth corner_freq magnitude mag_type
stat_code comp phase_ttime azimuth hypo_dist signal_moment blank t-star alpha corner_freq
stat_code comp phase_ttime azimuth hypo_dist signal_moment blank t-star alpha corner_freq
stat_code comp phase_ttime azimuth hypo_dist signal_moment blank t-star alpha corner_freq
.
.
.
stat_code comp phase_ttime azimuth hypo_dist signal_moment blank t-star corner_freq
# yyyymmdd hhmi ss.ss deg_lat min_lat deg_lon min_lon depth corner_freq magnitude mag_type
stat_code comp phase_ttime azimuth hypo_dist signal_moment blank t-star alpha corner_freq
.
.
.
```

Example format:

```
# 19920227 0250 24.90 55.21 0.00 -3.42 0.00 5.90 22.49 2.70 L
```

XAL_ S_GZ 25.17 116.91 86.54 5.09993e-08 0.020700 0.00 22.49
XDE_ S_GZ 23.34 183.33 79.07 7.34799e-08 0.043200 0.00 22.49
XSO_ S_GZ 23.04 67.21 80.57 8.87064e-08 0.013500 0.00 22.49
19920228 0050 27.40 53.33 0.00 -1.18 0.00 11.60 34.89 2.00 L
LHO_ S_GZ 15.10 298.49 52.42 2.76217e-08 0.017400 0.00 34.89
.
.
.

OUTPUT FILE FORMATS

smom_corr.out

Generic Format:
yyyymmdd hhmi ss.ss deg_lat deg_lon depth corner_freq computed_mag record_mag
stat_code phase_ttime real_smom model_smom fr_diff M3norm_smom hypo_dist corner_freq t-star alpha azimuth

Example format:
19921230 2134 13.00 47.71 8.38 22.00 7.2468 3.31 3.70
622_ 17.14 2.50149e-06 1.42842e-06 -0.42897 8.64950e-07 60.00 7.2468 2.366900e-02 0.000 0.00
693_ 17.43 7.26900e-06 1.40500e-06 -0.80671 2.51343e-06 61.00 7.2468 2.388500e-02 0.000 0.00
855_ 15.71 1.34288e-06 1.55827e-06 0.16040 4.64333e-07 55.00 7.2468 2.327100e-02 0.000 0.00
19930214 1017 45.00 37.71 21.38 4.00 1.2235 4.77 4.50
214_ 2.29 1.69481e-03 1.69481e-03 0.00000 3.70432e-06 8.00 1.2235 1.890000e-02 0.000 0.00
.
.
.

Xout.dat

Xout.dat is a list of all ln(plateau value), geometrical spreading exponents, and ln(site amplifications) in that order. The number of rows is therefore dependant on the nmber of events and stations used, however there are always 5 decay values. If not used in the inversion they will be show as 0 values. The first four values are the body wave exponents of decay, the first being the closest segment to the source, the last being the furthest. The fifth value is the surface wave exponent of decay.

plot_gspr.dat

List of distances (km) and corresponding signal moment of a M3 event using the geometrical spreading model found along with no station amplification. Useful for plotting the geometrical spreading function.

plot_surf_gspr.dat

As for plot_gspr.dat but for the surface wave model.

Sinv_out.dat

List of eigenvalues in order of magnitude.

A_out.dat

The kernal matrix (A): y = A.m; where y is the data vector and m is the model vector.

NEM_out.dat

The Normal Equations Matrix setup for the inversion.

NEM_INV_out.dat

The inverted Normal Equations Matrix.

COMMON EXAMPLES

1. sampfit moment.dat 100000/100000/100000 100000 -z
   Run with a single segment geometrical spreading model (1/R**x) using all data.
2. sampfit moment.dat 100000/100000/100000 100000 -z -n5/5 -D10/300 -M3/5 -A80
   Run with a single segment geometrical spreading model (1/R**x) using data from earthquakes with 5 or more records recorded on stations with a total of 5 or more records. The records must be recorded between 10 and 300km in hypocentral distance and the earthquake magnitude must be between 3 and 5. The azimuth coverage for each earthquake must be greater than 80 degrees in standard deviation for the event to be included.
3. sampfit moment.dat 30/50/100000 5 -z -n5/5 -D10/300 -M3/5 -A80
   As above (2) but run with a three segment geometrical spreading model (1/R**a for 0 to 30km; 1/R**b for 30 to 50km; and 1/R**c for above 50km). Also discard data 5 standard deviations away from the mean and reiterate.
4. sampfit moment.dat 30/50/150 5 -z -n5/5 -D10/300 -M3/5 -A80 -f1/0/1.5/1
   As above (3) but specify the geometrical spreading model exponents as 1, 0, 1.5 and 1 respectively: Note all exponents must be specified (you cannot specify two and invert the other two).
5. sampfit moment.dat 30/50/110 5 -z -n5/5 -F1.1
   Specify that the first segment of decay is 1/R**1.1 (0 to 30km). Invert for the remaining values.
6. sampfit moment.dat 100000/100000/100000 100000 -z -o1
   As (1) but forward calculate moments using 1 bar stress drop.
7. sampfit moment.dat 100000/100000/100000 100000 -z -o1 -f1
   As (6) but set geometrical spreading as 1/R: we are now just inverting for the station amplifications.
8. sampfit moment.dat 30/70/110 100000 -z -o1 -f1/0/1/1.5 -n1/0 -s150 -g0.5
   We specify a four segment geometrical spreading model for body waves and fix the decay values using -f. Records beyond 150km are treated as surface waves for which we fix their decay to 1/R**0.5 (we could leave -g out and invert for this instead).
9. sampfit moment.dat 30/70/110 100000 -z -e -V./mod.dat
   Segments one (0 to 30km) and four (110km+) share the same decay value. We are using a velocity model called "mod.dat".

SEE ALSO

sfreqfit(1), cspecmc(2),

Edwards, B., A. Rietbrock, J. J. Bommer, and B. Baptie (2008). The Acquisition of Source, Path and Site Effects from Micro-earthquake Recordings using Q Tomography: Application to the UK, submitted to Bull. Seismol. Soc. Am.

BUGS

Please email author.

AUTHOR

Benjamin Edwards - b.edwards@liv.ac.uk